
Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

srch_pp      protein - protein database search, using Smith-Waterman algorithm
n on:        Wed Sep 1 16:16:39 1999;      MasPar time 6.77 Seconds
                    550.316 Million cell updates/sec
bular output not generated.

```

```

title:
>PCT-US99-13024-2
description:
(1-93) from PCTUS9913024 pep (6 of 12)
634
irect Score:
quence:
1 MEKFMAEFGGQYVQTPFSE.....DSKSYAFSTNDTTSAAFSV 93

oring table:
PAM 150
Gap 11

```

arched: 122810 seqs, 40068593 residues

```

st-processing: Minimum Match 0%
                Listing first 45 summaries

```

tabase: pir60
1:pir1 2:pir2 3:pir3 4:pir4

statistics: Mean 38.761; Variance 72.705; scale 0.533

SUMMARIES

sult	No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	92	14.5	299	2	S61248	hypothetical protein	1.99e-01
2	92	14.5	521	1	HYBSN	baicilolysin (EC 3.4.	1.99e-01
3	89	14.0	566	2	G71525	probable OMP [leader	5.25e-01
4	86	13.6	529	2	A28091	licheninase (EC 3.2.1	1.35e+00
5	86	13.6	521	2	I39956	neutral proteinase (E	1.35e+00
6	86	13.6	574	2	S28375	hypothetical protein	1.35e+00
7	86	13.6	673	2	A45456	NADH dehydrogenase (u	1.35e+00
8	86	13.6	1693	1	MNWHE	genome polyprotein -	1.35e+00
9	85	13.4	242	1	LXBS	licheninase (EC 3.2.1	1.85e+00
10	85	13.4	331	2	A61046	ecdysone-induced memb	1.85e+00
11	85	13.4	869	2	C56617	cfac protein precursor	1.85e+00
12	84	13.2	243	2	S15388	licheninase (EC 3.2.1	2.52e+00
13	84	13.2	622	2	S17402	parasporal crystal pr	2.52e+00
14	83	13.1	262	2	D70961	probable echaI protein	3.42e+00
15	83	13.1	281	2	S38913	hypothetical protein	3.42e+00
16	81	13.1	611	2	D70928	hypothetical protein	3.42e+00
17	83	13.1	613	2	T00758	ethyline kinase sen	3.42e+00
18	82	12.9	81	2	I38375	tyrosine kinase - hum	4.63e+00
19	82	12.9	449	2	S55092	hypothetical protein	4.63e+00
20	82	12.9	486	2	F69762	transporter homolog y	4.63e+00
21	82	12.9	527	2	I84483	tyrosine kinase - hum	4.63e+00
22	81	12.8	216	2	G70447	flagellar L-ring prot	6.25e+00
23	81	12.8	345	2	S45767	probable GTP cycloh	6.25e+00

24	81	12.8	377	2	H71282	hypothetical protein	6.25e+00
25	81	12.8	433	2	H70141	enolase (eno) homolog	6.25e+00
26	81	12.8	527	2	I49133	Txk - mouse	6.25e+00
27	81	12.8	527	2	A55631	protein-tyrosine kina	6.25e+00
28	81	12.8	3011	2	A40770	polyprotein precursor	6.25e+00
29	80	12.6	171	2	B38162	hypothetical protein	8.42e+00
30	80	12.6	300	2	I39495	hypothetical protein 4	8.42e+00
31	80	12.6	572	2	S14200	GRESAG protein - Tryp	8.42e+00
32	80	12.6	584	2	S27500	xylanase - Prevotella	8.42e+00
33	80	12.6	802	2	S48529	NAB3 protein - yeast	8.42e+00
34	80	12.6	954	2	S46105	glucan 1,4-alpha-gluc	8.42e+00
35	80	12.6	1659	2	JC4956	vitellogenin precursor	8.42e+00
36	80	12.6	3011	1	GNWC3	genome polyprotein -	8.42e+00
37	79	12.5	370	2	S69718	hypothetical protein	1.13e+01
38	79	12.5	469	2	S53810	paracrystalline surfa	1.13e+01
39	79	12.5	474	2	S39787	vitamin D-binding pro	1.13e+01
40	79	12.5	616	2	A40595	methionyl-CoA mut	1.13e+01
41	79	12.5	633	2	D32053	parasporal crystal pr	1.13e+01
42	79	12.5	744	2	S5926	NADH dehydrogenase (u	1.13e+01
43	79	12.5	744	2	S71864	NADH dehydrogenase (u	1.13e+01
44	79	12.5	860	2	F71000	hypothetical protein	1.13e+01
45	79	12.5	935	2	S57080	hypothetical protein	1.13e+01

ALIGNMENTS

RESULT 1

	ENTRY
	TITLE
	ORGANISM
	DATE

ACCESSIONS
REFERENCE
#authors

```
#submission
#description
```

```
#accession
##status
##molecule
```

##residue
##cross-r
CLASSIFICATION
SUMMARY

Query Match	
Best Local Sim	17
Matches	

Db	45	PREVC
		:
Qv	41	PREVC

RESULT 2

ENTRY
TITLE

ALTERNATE_NAMES	ORGANISM

ACCESSIONS
REFERENCE

#journal
#title

```

S61248      #type complete
             hypothetical protein - bovine herpesvirus 1
             #formal_name bovine herpesvirus 1
             18-Sep-1997 #sequence_revision 18-Sep-1997 #text_change
                   21-Aug-1998
S61248
S61233      D.; Letchworth, G.J.; Schwyzler, M.
             Vlecek, C.; Benes, V.; Lu, Z.; Kutish, G.F.; Paces, V.; Rock,
             submitted to the EMBL data Library, January 1995
             Nucleotide sequence analysis of a 30-kb region of the bovine
             herpesvirus 1 genome which exhibits a collinear gene
             arrangement with the UL21 to UL4 genes of herpes simplex
             virus.
S61248
             preliminary
             type DNA
                   1-299 ##label VLC
             references EMBL:Z48053; NID:g971311; PID:g971327
             ##superfamily varicella-zoster virus gene 53 protein
             #length 299 #molecular_weight 32278 #checksum 1192

```

```

Query Match      14.5%  Score 92;  DB 2;  Length 299;
Best Local Similarity 47.2%  pred. No. 1.99e-01;
Matches 17;  Conservative 9;  Mismatches 6;  Indels 4;  Gaps 4;

```

RESULT	2
ENTRY	
TITLE	HYB5N #type complete bacillolysin (EC 3.4.24.28) precursor - Bacillus amylioliquefaciens
ALTERNATE_NAMES	Bacillus metalloendopeptidase; microbial metalloproteinase; neutral proteinase
ORGANISM	#formal_name Bacillus amylioliquefaciens
DATE	30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change

05-Sep-1997
ACCESSIONS
A25415
REFERENCE
A25415
#authors
Vasantha, N.; Thompson, L.D.; Rhodes, C.; Banner, C.; Nagle, J.; Filiputa, D.
#journal
J. Bacteriol. (1984) 159:811-819
#title
Genes for alkaline protease and neutral protease from

```

Bacillus amyloliquefaciens contain a large open reading
frame between the regions coding for signal sequence and
mature protein.
#cross-references MUID:85006739
#accession A25415
##molecule_type DNA
##residues 1-521 ##label YAN
##cross-references GB:K02497; NID:g143248; PID:g143249
GENETICS
#gene npr
#start_codon GTG
CLASSIFICATION #superfamily thermolysin
KEYWORDS calcium; extracellular protein; hydrolase; metalloproteinase;
zinc
FEATURE
1-27
28-221 #domain signal sequence #status predicted #label SIG\
222-521 #domain propeptide #status predicted #label PRO\
#product bacillolysin #status predicted #label MAT\
#binding_site zinc (His, His, Glu) #status predicted\
#active_site Glu, His #status predicted
#length 521 #molecular-weight 56840 #checksum 7195
SUMMARY
Query Match 14.5%; Score 92; DB 1; Length 521;
Best Local Similarity 27.8%; Pred. No. 1.99e-01;
Matches 15; Conservative 19; Mismatches 18; Indels 2; Gaps 2;
Db 240 ISSSEKGYVLRLSKPTGTQIIITYDLQNRVNLPGTLVSTTNQFTSSORAA 292
Qy 36 LSTAGPSYV-KFQDNPVGSOTFSAGLHLRVDPSTGALVDSKSYAFSTNDTTS 88

RESULT 3
ENTRY G71525 #type complete
TITLE Probable OMP [leader (19) peptide] - Chlamydia trachomatis
(seroType D, strain OW3/Cx)
ORGANISM #formal_name Chlamydia trachomatis
DATE 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change
21-Nov-1998
ACCESSIONS G71525
REFERENCE A71570
#authors Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe,
R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov,
R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
#journal Science (1998) 282:754-759
#title Genome sequence of an obligate intracellular pathogen of
humans: Chlamydia trachomatis.
#cross-references MUID:99000809
#accession G71525
##status preliminary
##molecule_type DNA
##residues 1-566 ##label ARN
##cross-references GB:AE001308; GB:AE001273; NID:g3328766; PID:g3328771
##experimental_source serotype D, strain UW-3/Cx
GENETICS
#gene CT350
SUMMARY
#length 566 #molecular-weight 63507 #checksum 4960
Query Match 14.0%; Score 89; DB 2; Length 566;
Best Local Similarity 26.2%; Pred. No. 5.25e-01;
Matches 17; Conservative 25; Mismatches 18; Indels 5; Gaps 5;
Db 63 IAESYLOQSFLEDTYIR-KSAITGA-GLGSSEA-LELLSEATETQDLYEQ-L-ILNA 117
Qy 8 FGQGYVTPFLESNSVRYKISAGSCPLSTAGPSYVKFQDNPVGSOTFSAGLHLRVDP 67
Db 118 ATSL 122
Qy 68 STGAL 72

RESULT 4
ENTRY A29091 #type complete
TITLE licheninase (EC 3.2.1.73) beta - Bacillus amyloliquefaciens

```

```

ALTERNATE_NAMES 1.3-1.4-beta-D-glucan 4-glucanohydrolase; beta-glucanase;
lichenase
ORGANISM #formal_name Bacillus amyloliquefaciens
DATE 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change
20-Mar-1998
ACCESSIONS A29091
REFERENCE A91564
#authors Hofemaister, J.; Kurtz, A.; Borriess, R.; Knowles, J.
#journal Gene (1986) 49:177-187
#title The beta-glucanase gene from Bacillus amyloliquefaciens shows
extensive homology with that of Bacillus subtilis.
#cross-references MUID:87192007
#accession A29091
##molecule_type DNA
##residues 1-239 ##label HOF
##cross-references GB:M15674; NID:g143009; PID:g143010
##experimental_source strain BE20/78
GENETICS
#gene bglA
CLASSIFICATION #superfamily licheninase
KEYWORDS glycosidase; hydrolase; polysaccharide degradation
SUMMARY #length 239 #molecular-weight 26928 #checksum 1611
Query Match 13.6%; Score 86; DB 2; Length 239;
Best Local Similarity 25.5%; Pred. No. 1.35e+00;
Matches 14; Conservative 20; Mismatches 18; Indels 3; Gaps 3;
Db 74 LAUTSPYKFDGCGNRSVQTYGYLYEVRM-KPAKNTGIVSSFFTYGTGTEGP 127
Qy 36 LSTAGPSYV-KFQDNPVGSOTFSAGLHLRVDPSTGALVDSKSYAFSTNDTTS 88

RESULT 5
ENTRY I39956 #type complete
TITLE neutral proteinase (EC 3.4.24.-) - Bacillus amyloliquefaciens
ORGANISM #formal_name Bacillus amyloliquefaciens
DATE 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change
18-Mar-1997
ACCESSIONS I39956
REFERENCE I39956
#authors Shimada, H.; Honjo, M.; Mita, I.; Nakayama, A.; Akaoka, A.;
Manabe, K.; Furutani, Y.
#journal J. Biotechnol. (1985) 2:75-85
#title The nucleotide sequence and some properties of the neutral
protease gene of Bacillus amyloliquefaciens.
#accession I39956
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-521 ##label RES
##cross-references GB:M36723; NID:g143352; PID:g143353
GENETICS
#start_codon GTG
CLASSIFICATION #superfamily thermolysin
KEYWORDS hydrolase; metalloproteinase
SUMMARY #length 521 #molecular-weight 56725 #checksum 6816
Query Match 13.6%; Score 86; DB 2; Length 521;
Best Local Similarity 27.8%; Pred. No. 1.35e+00;
Matches 15; Conservative 18; Mismatches 19; Indels 2; Gaps 2;
Db 240 ISSSEKGYVLRLSKPTGTQIIITYDLQNRVNLPGTLVSTTNQFTSSORAA 292
Qy 36 LSTAGPSYV-KFQDNPVGSOTFSAGLHLRVDPSTGALVDSKSYAFSTNDTTS 88

RESULT 6
ENTRY S28275 #type fragment
TITLE hypothetical protein F54G8.4 - Caenorhabditis elegans
(fragment)
ORGANISM #formal_name Caenorhabditis elegans
DATE 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change
09-Sep-1997
ACCESSIONS S28275

```


[illegible]

```

Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Dueterhoeft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelie, D.; Porwollik, S.; Prescott,
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Scanlon, E.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession B69594
#status preliminary: nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-242 #label KUN
#cross-references GB:295124; GB:AL009126; NID:g2636442; PID:el184632;
#experimental_source PID:g2636453
#experimental_source strain 168
REFERENCE
A93526
Murphy, N.; McConnell, D.J.; Cantwell, B.A.
Nucleic Acids Res. (1984) 12:5355-5367
#journal
#title The DNA sequence of the gene and genetic control sites for
the excreted B. subtilis enzyme beta-glucanase.
#cross-references MUID:8427222
#accession A22914
#molecule_type DNA
#residues 1-203, 'L', 205-242 #label MUR
#cross-references EMBL:X00754; NID:g39818; PID:g685236
#experimental_source strain C120
#note the authors translated the codon CAA for residue 29 as
Lys and CCA for residue 82 as Leu
REFERENCE
A90026
Tezuka, H.; Yuuki, T.; Yabuuchi, S.
Agric. Biol. Chem. (1989) 53:2335-2339
#journal
#title Construction of a beta-glucanase hyperproducing Bacillus
subtilis using the cloned beta-glucanase gene and a
multi-copy plasmid.
#accession A90026
#molecule_type DNA
#residues 1-23, 'S', 25-82, 'S', 84-242 #label TEZ
#cross-references DBAJ:D00518; NID:g216243; PID:g216244
#experimental_source strain Y-25, clone pLE100
REFERENCE
A90027
Yuuki, T.; Tezuka, H.; Yabuuchi, S.
Agric. Biol. Chem. (1989) 53:2341-2346
#journal
#title Purification and some properties of two enzymes from a
beta-glucanase hyperproducing strain, Bacillus subtilis
HL-25.

```

```

#contents
#note source was hyperproducing strain HL-25 with gene from strain
Y-25
#note the amino ends of the mature forms of E-1 and E-2 are
pyroglutamic acid and glutamine, respectively
GENETICS
#gene bglis
FUNCTION
#description catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in
lichenin and other 1,3- and 1,4-beta-linked polysaccharides
#superfamily licheninase
CLASSIFICATION
#extracellular protein; glycosidase: hydrolase; polysaccharide
degradation; pyroglutamic acid
KEYWORDS
FEATURE
1-28 #domain signal sequence #status predicted #label SIG\
29-242 #product licheninase #status predicted #label MAT\
29 #modified_site pyrrolidone carboxylic acid (Gln) (in
mature form) (partial) #status experimental
SUMMARY
#length 242 #molecular-weight 27268 #checksum 4439
Query Match 13.4%; Score 85; DB 1; Length 242;
Best Local Similarity 25.5%; Pred. No. 1.85e+00;
Matches 14; Conservative 20; Mismatches 18; Indels 3; Gaps 3;
Db 77 LALTSPAYNFKDCGENRSVQTYGYLYEVRM-KPAKNTGIVSSFTYTGTGTP 130
I:::I::I::: : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 36 LSTAGPSYVVKFQ-DNPVGSQTFNSAGLH-LRVFDPSTGALVDSKSYAFSTNDTTS 88
I:::I::I::: : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 10
ENTRY #type complete
TITLE ecdysone-induced membrane protein IMP-E3 - fruit fly
ORGANISM (Drosophila melanogaster)
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
16-Feb-1997
ACCESSIONS
REFERENCE A61046
#authors Moore, J.T.; Fristrom, D.; Hammonds, A.S.; Fristrom, J.W.
#journal Dev. Genet. (1990) 11:299-309
#title Characterization of IMP-E3, a gene active during imaginal
disc morphogenesis in Drosophila melanogaster.
#accession A61046
#status preliminary
#molecule_type mRNA
#residues 1-331 #label MOO
GENETICS
#gene FlyBase:ImpE3
#cross-references FlyBase:FBgn0001255
KEYWORDS membrane protein
SUMMARY #length 331 #molecular-weight 36583 #checksum 8221
Query Match 13.4%; Score 85; DB 2; Length 331;
Best Local Similarity 31.9%; Pred. No. 1.85e+00;
Matches 15; Conservative 12; Mismatches 18; Indels 2; Gaps 2;
Db 190 LDNFLRLYDNYGPAFGSESAMDWRSTAGTKRKYPTKP-VVDF 235
QY 1 MEKFMFEGGYVQTFPLESSESVRYKI-STAGSCPLSTAGPSYVKF 46
I:::I::: : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 11
ENTRY #type complete
TITLE cfsc protein precursor - Escherichia coli plasmid NTP113
ORGANISM #formal_name Escherichia coli
DATE 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change
20-Mar-1998
ACCESSIONS
REFERENCE C56617
#authors Jordi, B.J.; Willshaw, G.A.; van der Zeijst, B.A.; Gaastra,
W.
#journal DNA Seq. (1992) 2:257-263
#title The complete nucleotide sequence of region 1 of the CFA/I

```

```

fimbrial operon of human enterotoxigenic Escherichia coli.
#cross-references MUID:92329981
#accession C56617
#status preliminary
#molecule_type DNA
#residues 1-869 #label JOR
#cross-references GB:M5661; MID:g145507; PID:g145510
#experimental_source enterotoxigenic strain, CFA/I-ST plasmid NTP113
#note sequence extracted from NCBI backbone (NCBI:108960,
NCBIP:108971)

GENETICS
#gene cfaC
#genome plasmid
SUMMARY
#length 869 #molecular-weight 97830 #checksum 9755

Query Match 13.4%; Score 85; DB 2; Length 869;
Best Local Similarity 32.4%; Pred. No. 1.85e+00;
Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;

Db 158 AFISQITINLSDSKYKRLISGNSALGITDTSY 191
::: : : : : : : : : : : : : : : : : :
Qy 11 GYVQTPFLSESNVRYK-ISIAGSCLSTAGPSY 43

RESULT 12
ENTRY #type complete
TITLE licheninase (EC 3.2.1.73) - Bacillus licheniformis
ALTERNATE_NAMES endo-beta-1,3-1,4-D-glucanase; lichenase
ORGANISM #formal name Bacillus licheniformis
DATE 21-Nov-1993 #sequence_revision 01-Dec-1995 #text_change
08-Sep-1997
ACCESSIONS S15388
REFERENCE S15388
#authors Lloberas, J.; Perez-Pons, J.A.; Querol, E.
#journal Eur. J. Biochem. (1991) 197:337-343
#title Molecular cloning, expression and nucleotide sequence of the
endo-beta-1,3-1,4-D-glucanase gene from Bacillus
licheniformis. Predictive structural analyses of the
encoded polypeptide.
#cross-references MUID:91224124
#accession S15388
#status preliminary
#molecule_type DNA
#residues 1-243 #label LLO
#cross-references EMBL:X57279; NID:G39558; PID:G39559
CLASSIFICATION #superfamily licheninase
KEYWORDS glycosidase; hydrolase; polysaccharide degradation
SUMMARY #length 243 #molecular-weight 27435 #checksum 9335

Query Match 13.2%; Score 84; DB 2; Length 243;
Best Local Similarity 27.8%; Pred. No. 2.52e+00;
Matches 15; Conservative 15; Mismatches 23; Indels 1; Gaps 1;

Db 78 LSLTSPSYNFKDGENRSVOTGYGLYEVNMPKPNKNGVIVSSFFTYTGPDPGP 131
::: : : : : : : : : : : : : : : : : :
Qy 36 LSTAGPSYVRFQ-ONPYGQTSFAGLHRLVDFDPSTGALVDKSYAFSTSDN 88

RESULT 13
ENTRY #type complete
TITLE parasporal crystal protein cryIIC - Bacillus thuringiensis
plasmid
ALTERNATE_NAMES delta-endotoxin
ORGANISM #formal name Bacillus thuringiensis
DATE 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
09-Sep-1997
ACCESSIONS S17402
REFERENCE S17402
#authors Wu, D.; Cao, X.L.; Bai, Y.Y.; Aronson, A.I.
#journal FEMS Microbiol. Lett. (1991) 81:31-36
#title Sequence of an operon containing a novel delta-endotoxin gen
from Bacillus thuringiensis.
#accession S17402

```


工 口 司 同 向

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.

544.173 Million cell updates/sec

Result	No.	Score	Query		Length	DB	ID	Description	Pred. No.
			Match						
1	300	47.3	65	1	V15A_BP74		HYPOTHEICAL 7.3 KD PR	1.60e-44	
2	100	15.8	1693	1	POLN_HEVWY		NON-STRUCTURAL POLYPRO	3.02e-03	
3	92	14.5	521	1	NPRE_BACAM		BACILLOLYSIN PRECURSOR	5.71e-02	
4	87	13.7	703	1	G411_RHISN		HYPOTHEICAL 76.2 KD P	3.31e-01	
5	86	13.6	239	1	YUB_BACAM		BETA-GLUCANASE PRECURS	4.66e-01	
6	86	13.6	457	1	EMB8_PIGCL		LATE EMBRYOGENESIS ABU	4.66e-01	
7	86	13.6	672	1	NQO3_PARDE		NADH-UBIQUINONE OXIDOR	4.66e-01	
8	86	13.6	932	1	YMB4_CAEEL		HYPOTHEICAL 104.4 KD	4.66e-01	
9	86	13.6	1693	1	POLN_HEVBU		NON-STRUCTURAL POLYPRO	4.66e-01	
10	86	13.6	1693	1	POLN_HEVPA		NON-STRUCTURAL POLYPRO	4.66e-01	
11	85	13.4	242	1	GUB_BACSU		BETA-GLUCANASE PRECURS	6.55e-01	
12	85	13.4	869	1	CFAC_ECOLI		CFA/I FIMBRIAL SUBUNIT	6.55e-01	
13	84	13.2	243	1	GUB_BACLI1		BETA-GLUCANASE PRECURS	9.18e-01	
14	83	13.1	611	1	YX45_MYCTU		HYPOTHEICAL 67.2 KD P	1.28e+00	
15	82	12.9	449	1	YM60_YEAST		HYPOTHEICAL 51.4 KD P	1.79e+00	
16	82	12.9	527	1	TXK_HUMAN		TYROSINE-PROTEIN KINAS	1.79e+00	
17	81	12.8	345	1	GCH2_YEAST		GTP CYCLOHYDROLASE II	2.48e+00	
18	81	12.8	433	1	ENO_BORBU		ENOLASE (EC 4.2.1.11)	2.48e+00	
19	81	12.8	527	1	TXK_MOUSE		TYROSINE-PROTEIN KINAS	2.48e+00	
20	80	12.6	171	1	YCB2_PSEDE		HYPOTHEICAL 19.0 KD P	3.43e+00	
21	80	12.6	300	1	NTRY_AZOBR		NITROGEN REGULATION PR	3.43e+00	
22	80	12.6	502	1	CI42_TRYBB		RECEPTOR-TYPE ADENYLAT	3.43e+00	
23	80	12.6	872	1	NAB3_YEAST		NUCLEAR POLYAADENYLAT	3.43e+00	

```
CC ----- TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH). -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch). -----
CC
CC EMBL: D10330; G221705; ..
CC POLYPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; HELICASE;
CC ATP-BINDING. 975 982 ATP (POTENTIAL).
CC NP_BIND 975 982 ATP (POTENTIAL).
CC SEQUENCE 1693 AA; 185215 MW; FFCB786D CRC32;
CC
CC Query Match 15.8%; Score 100; DB 1; Length 1693;
CC Best Local Similarity 26.9%; Pred. No. 3.02e-03;
CC Matches 18; Conservative 22; Mismatches 24; Indels 3; Gaps 3;
CC
CC Db 572 FRTSFGDVGAVLEANGPERYNLSFDASQSTMAAGPFSLTLYAASAGLEVVYAAGLDHRAV 631
CC | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
CC Qy 8 FGQGVQTPFLSESNVRYKISAGSCPLSTAGPSYVKFQDNPVGSQT-F-SAGLHLR-V 64
CC
CC Db 632 FAPGVSP 638
CC | : : :
CC Qy 65 FDPSTGA 71
CC
CC RESULT 3
CC ID NPBE_BACAM STANDARD; PRT; 521 AA.
CC AC P06832;
CC DT 01-JAN-1988 (REL. 06, CREATED)
CC DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
CC DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
CC DE BACILLOLYSIN PRECURSOR (EC 3.4.24.28) (NEUTRAL PROTEASE).
CC GN NPR
CC OS BACILLUS AMYLOLIQUEFACIENS.
CC OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
CC OC BACILLUS.
CC [1]
CC RN SEQUENCE FROM N.A.
CC RP STRAIN-ATCC 23844;
CC RC MEDLINE; 85006739.
CC RA VASANTHA N., THOMPSON L.D., RHODES C., BANNER C., NAGLE J.,
CC RA FILPUA D.;
CC RT "Genes for alkaline protease and neutral protease from Bacillus
CC RT amyloliquefaciens contain a large open reading frame between the
CC RT regions coding for signal sequence and mature protein.";
CC RL J. BACTERIOL. 159:811-819(1984).
CC CC -!- FUNCTION: THERMOLABILE EXTRACELLULAR ZINC METALLOPROTEASE.
CC CC -!- CATALYTIC ACTIVITY: SIMILAR, BUT NOT IDENTICAL, TO THAT OF
CC CC THERMOLYSIN.
CC CC -!- COFACTOR: BINDS AND REQUIRES A ZINC ATOM, WHICH IS ESSENTIAL FOR
CC CC PROTEOLYTIC ACTIVITY. BINDS FOUR CALCIUM IONS (BY SIMILARITY).
CC CC -!- SUBCELLULAR LOCATION: SECRETED.
CC CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M4 (ZINC METALLOPROTEASE);
CC CC ALSO KNOWN AS THE THERMOLYSIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch). -----
CC
CC EMBL: K02497; G143249; ..
CC PIR: A25415; HYBSN.
CC DR PROSITE: PS00142; ZINC_PROTEASE; 1.
CC DR PFAM: PF00099; zn-protease; 1.
CC DR HSSP: P00800; ITRL.
CC KW HYDROLASE; METALLOPROTEASE; ZINC; CALCIUM; ZYMOGEN; SIGNAL.
```

```
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 221 ACTIVATION PEPTIDE.
FT CHAIN 222 521 BACILLOLYSIN.
FT METAL 364 364 ZINC (CATALYTIC) (BY SIMILARITY)..
FT ACT_SITE 365 365 BY SIMILARITY..
FT METAL 368 368 ZINC (CATALYTIC) (BY SIMILARITY)..
FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY)..
FT ACT_SITE 449 449 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 521 AA; 56840 MW; BC0147D4 CRC32;
CC
CC Query Match 14.5%; Score 92; DB 1; Length 521;
CC Best Local Similarity 27.8%; Pred. No. 5.71e-02;
CC Matches 15; Conservative 19; Mismatches 18; Indels 2; Gaps 2;
CC
CC Db 240 ISSSGKYLRLDLKPTGTOIITDYDLQNLREYNLP-GLVSVSTTNOFTTSSORAA 292
CC | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
CC Qy 36 LSTAGPSYV-KFQDNPVGSQTFSGALHLRVPDSTGALVDKSYAFSTSDTTTS 88
CC
CC RESULT 4
CC ID Y4II_RHISN STANDARD; PRT; 703 AA.
CC AC P55492;
CC DT 01-NOV-1997 (REL. 35, CREATED)
CC DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
CC DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
CC DE HYPOTHETICAL 76.2 KD PROTEIN Y4II.
CC GN Y4II.
CC OS RHIZOBIUM SP. (STRAIN NGR234).
CC OG PLASMID SYN PNR234A.
CC OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
CC OC RHIZOBIACEAE; RHIZOBIUM.
CC [1]
CC RN SEQUENCE FROM N.A.
CC RP MEDLINE; 97305956.
CC RX FREIBERG C.A., FELLAY R., BAIRDOCH A., BROUGHTON W.J., ROSENTHAL A.,
CC RA PERRET X.;
CC RT "Molecular basis of symbiosis between Rhizobium and legumes.";
CC RL NATURE 387:394-401(1997).
CC CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC CC -!- SIMILARITY: NONE OBVIOUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch). -----
CC
CC EMBL: AE000078; G2182446; ..
CC DR HYPOTHETICAL PROTEIN; PLASMID; TRANSMEMBRANE.
CC KW TRANSNEM 23 43 POTENTIAL.
CC FT TRANSNEM 69 89 POTENTIAL.
CC FT TRANSNEM 143 163 POTENTIAL.
CC FT TRANSNEM 250 270 POTENTIAL.
CC FT TRANSNEM 357 377 POTENTIAL.
CC FT TRANSNEM 432 452 POTENTIAL.
CC FT TRANSNEM 644 664 POTENTIAL.
CC SEQUENCE 703 AA; 76183 MW; A2BA53CE CRC32;
CC
CC Query Match 13.7%; Score 87; DB 1; Length 703;
CC Best Local Similarity 26.5%; Pred. No. 3.31e-01;
CC Matches 18; Conservative 20; Mismatches 28; Indels 2; Gaps 2;
CC
CC Db 224 SPFADTDSGTISATSLFSGAGASTRPPWLAFAANAALASGAPFPGISPRVIATATSOYA 283
CC | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
CC Qy 15 TPLF-SESNVRYKISAGSCPLSTAGPSYVKFQDNPVGSQTFSGALHLRVPDSTGALV 73
CC
CC Db 284 KAK-FPIS 290
CC | : : :
CC Qy 74 DSKSYAFS 81
```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; L47118; GI350545; -.
DR PROSITE; PS01133; UPF0017.1.
SQ SEQUENCE 457 AA; 51019 MW; AE7CB4CD CRC32;

Query Match 13.6%; Score 86; DB 1; Length 457;
Best Local Similarity 34.0%; Pred. No. 4.66e-01;
Matches 17; Conservative 11; Mismatches 20; Indels 2; Gaps 2;

Dd 232 LGANILRYLGEVAGNCPLESAVSLCNPPF-NLVIAEDFHKGGLGFNNVYD 280
| : | ||| : ||||| : | : : | || : | : |
Qy 18 LSENSRVYKISAIAGSCPLSTAGESYVKFDNPVPVGSQTFSAGLHLR-VFD 66

RESULT 7

ID NOO3 PARDE STANDARD; PRT; 672 AA.

AC A29915;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 3 (EC 1.6.5.3) (NADH
DE DEHYDROGENASE 1, CHAIN 3) (NDH-1, CHAIN 3).
GN NOO3.
OS PARACOCCLUS DENITRIFICANS.
OS BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHODOBACTER GROUP;
RC PARACOCCLUS.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.
RX STRAIN=ATCC 13348;
RK MEDLINE; 92296779.
RL XU X., MATSUONO-YAGI A., YAGI T.;
RT "Structural features of the 66-kDa subunit of the energy-transducing
RT NADH-ubiquinone oxidoreductase (NDH-1) of Paracoccus denitrificans.";
RL ARCH. BIOCHEM. BIOPHYS. 296:40-48(1992).
RN [2]

RX SEQUENCE OF 657-672 FROM N.A.
RK MEDLINE; 93136200.
RL XU X., MATSUONO-YAGI A., YAGI T.;
RT "DNA sequencing of the seven remaining structural genes of the gene
RT cluster encoding the energy-transducing NADH-quinone oxidoreductase
RT of Paracoccus denitrificans";
RL BIOCHEMISTRY 32:968-981(1993).
CC -|- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC -|- COFACTOR: MAY BIND TWO 4FE-4S CLUSTER AND ONE 2FE-2S CLUSTER.
CC -|- SUBUNIT: COMPOSED OF 14 DIFFERENT SUBUNITS.
CC -|- SIMILARITY: BELONGS TO THE COMPLEX I 75 KD SUBUNIT FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; M84572; G150601; -.
DR PIR; S23948; S23948.
DR PROSITE; PS00641; COMPLEX1_75K_1; 1.
DR PROSITE; PS00642; COMPLEX1_75K_2; 1.
DR PROSITE; PS00643; COMPLEX1_75K_3; 1.
DR PFAM; PF00111; fer2; 1.
DR PFAM; PF00384; molybdopterin; 1.
KW OXIDOREDUCTASE; NAD; UBIQUINONE; IRON-SULFUR; 4FE-4S.
FT INIT_MET 0
FT METAL 25 25 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT METAL 36 36 IRON-SULFUR (2FE-2S) (POTENTIAL).

Query Match 13.6%; Score 86; DB 1; Length 932;
Best Local Similarity 38.2%; Pred. No. 4.66e-01;
Matches 13; Conservative 9; Mismatches 10; Indels 2; Gaps 2;

```

Db      282  TTFTEGLEVLNRFQADADELFNRK-LRFSAGDDAA 314
      ||: || || ||: | | | | | | | | | | | | | | | |
QY      55  TFSAGHL-LRVFDPSTGALVDSKSYAFSTSDTT 87

RESULT 9
ID      POLN_HEVBU      STANDARD;      PRT: 1693 AA.
AC      P29324;
DT      01-DEC-1992 (REL. 24, CREATED)
DT      01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT      15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE      NON-STRUCTURAL POLYPROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE
DE      (EC 2.7.7.48); HELICASE].
OS      HEPATITIS E VIRUS (STRAIN BURMA) (HEV).
OS      VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALICIVIRIDAE;
OC      CALICIVIRUS.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 92024667.
RA      TAM A.W., SMITH M.M., GUERRA M.E., HUANG C.-C., BRADLEY D.W.,
RA      FRY K.E., REYES G.R.;
RT      "Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT      full-length viral genome.";
RL      VIROLOGY 185:120-131(1991).
CC      -I- HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF ENTERICALLY
CC      TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; M73218; G330024; -
DR      PIR; A40778; MNWHE
KW      POLYPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; HELICASE;
KW      ATP-BINDING.
FT      NP_BIND 975 982 ATP (POTENTIAL).
SQ      SEQUENCE 1693 AA; 185191 MW; C560BE14 CRC32;

Query Match 13.68; Score 86; DB 1; Length 1693;
Best Local Similarity 25.48; Pred. NO. 4.66e-01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db      572  FRTSFVDGAVLENGPGRHNLSFDASOSTMAAGPFSLTAAASAGLEVRVVAAGLDHRAV 631
      | : | : | | : | : | : | : | : | : | : | : | : |
QY      8  FGQGYVQTPELSENSRVYKISITAGSCPLSTAGPSYVKFQDNVGSOT-F-SAGLHLR-V 64

Db      632  FAPGVSP 638
      | | : :
QY      65  FDPSTGA 71

RESULT 10
ID      POLN_HEVPA      STANDARD;      PRT: 1693 AA.
AC      P33424;
DT      01-FEB-1994 (REL. 28, CREATED)
DT      15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT      15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE      NON-STRUCTURAL POLYPROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE
DE      (EC 2.7.7.48); HELICASE].
OS      HEPATITIS E VIRUS (STRAIN PAKISTAN) (HEV).
OS      VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALICIVIRIDAE;
OC      CALICIVIRUS.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 92115700.
RA      TSAREV S.A., EMERSON S.U., REYES G.R., TSAREVA T.S., LEGTERS L.J.,
RA      MALIK I.A., IQBAL M., PURCELL R.H.;
RT      "Characterization of a prototype strain of hepatitis E virus.";

```


PLASMID NTP513
BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
ESCHERICHIA.
[1]
SEQUENCE FROM N.A.
STRAIN-ENTEROTOXIGENIC;
MEDLINE: 8930163.
HAMERS A.M., PEL H.J., WILLSHAW G.A., KUSTERS J.G.,
VAN DER ZEIJST B.A.M., GAASTRA W.;
"The nucleotide sequence of the first two genes of the CFA/I fimbrial operon of human enterotoxigenic Escherichia coli."; MICROB. PATHOG. 6:297-309(1989).
[2]
SEQUENCE FROM N.A.
MEDLINE: 92329981.
JORDI B.J.A.M., WILLSHAW G.A., VAN DER ZEIJST B.A.M., GAASTRA W.;
"The complete nucleotide sequence of region I of the CFA/I fimbrial DNA SEQ. 2:257-263(1992).
-!- FUNCTION: MAY SERVE AS ANCHOR FOR THE FIMBRAE IN THE OUTER MEMBRANE."

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL: M55661; GI45510; -
ANTIGEN; SIGNAL; FIMBRIA; OUTER MEMBRANE; PLASMID.
SIGNAL 1 22 POTENTIAL.
CHAIN 23 869 CFA/I FIMBRIAL SUBUNIT C.
SEQUENCE 869 AA: 97830 MW; 7AF6347 CRC32;

Query Watch 13.4%; Score 85; DB 1; Length 869;
Best Local Similarity 32.4%; Pred. No. 6.35e-01;
Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;

Db 158 AFISQITINLSGKYKRLSGNSALGITDTSY 191
::: : : 1: 1: 1: 1: 1: 1: : : : :
QY 11 GYVQTFFLESNSRVYK-TSIAGCPLSTAGPSY 43

RESULT 13
ID GUB_BACLI STANDARD; PRT; 243 AA.
AC P27051;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
GN (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
GD BGI.
OS BACILLUS LICHENIFORMIS.
OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RX SEQUENCE FROM N.A.
RA MEDLINE: 91224124.
RA LLOBERAS J., PEREZ-PONS J.A., QUEROL E.;
RT "Molecular cloning, expression and nucleotide sequence of the
RT endo-beta-1,3-1,4-D-glucanase gene from Bacillus licheniformis.
RT Predictive structural analyses of the encoded polypeptide.";
RL EUR. J. BIOCHEM. 197:337-343(1991).
RN [2]
RP REVISIONS.
RA QUEROL E.;
RL SUBMITTED (JUL-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP MUTAGENESIS.
RX MEDLINE: 92362869.
RA PLANAS A., JUNCOSA M., LLOBERAS J., QUEROL E.;

[illegible]

Search completed: Wed Sep 1 16:15:36 1999
Job time : 13 secs.

[illegible]

RP	SEQUENCE FROM N.A.					
RA	HAWKINS T., AINSCOUGH R.;					
RL	SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.					
CC	-1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.					
DR	EWBL; 22177; EI351235;					
DR	EWBL; 219155; EI351235; JOINED.					
DR	PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.					
KW	ZINC-FINGER.					
SQ	SEQUENCE 974 AA; 109137 MW; 4D7558C4 CRC32;					
Query Match 13.6%; Score 86; DB 5; Length 974;						
Best Local Similarity 38.2%; Pred. No. 1.51e+00;						
Matches 13; Conservative 9; Mismatches 10; Indels 2; Gaps 2;						
Dd	282 TFTGTEGVLEVNFPQDADELFNRK-LRESAGDDAA 314					
	: :: : : :: :					
Qy	55 TFSAGHLR-LRVFDPSTCALVDKSXYAFSTSDNTT 87					
RESULT	13					
ID	O69410 PRELIMINARY; PRT; 1693 AA.					
AC	O69410;					
DT	01-NOV-1996 (TREMBLREL. 01, CREATED)					
DT	01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)					
DT	01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)					
OS	METHYL TRANSFERASE.					
DE	HEPATITIS E VIRUS (HEV).					
OC	VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALICIVIRIDAE;					
OC	CALICIVIRUS.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=HEV037;					
RL	DONATI M.C., FAGAN E.A., HARRISON T.J.;					
RL	SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.					
DR	EWBL; X98292; EI175751;					
KW	TRANSFERASE.					
SQ	SEQUENCE 1693 AA; 185190 MW; 287742F6 CRC32;					
Query Match 13.6%; Score 86; DB 14; Length 1693;						
Best Local Similarity 25.4%; Pred. No. 1.51e+00;						
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;						
Dd	572 FRTSFVDGAVLETNGPERRHNLSFDASOSTMAAGPFSLTYAASAGLEVRYVAAGLDHRAV 631					
	:: : : :: : : :: : : :: : : :: :					
Qy	8 FGQGYVTPLSESNVYRIKISGTAGCSPLTAGPSYVKFODNPVGSOT-F-SAGLHLR-V 64					
Dd	632 FAGVSP 638					
	: :					
Qy	65 FDPSTGA 71					
RESULT	14					
ID	Q89444 PRELIMINARY; PRT; 1693 AA.					
AC	Q89444;					
DT	01-NOV-1996 (TREMBLREL. 01, CREATED)					
DT	01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)					
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)					
DE	UNNAMED PROTEIN PRODUCT.					
OS	HEPATITIS E VIRUS (HEV).					
OC	VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALICIVIRIDAE;					
OC	CALICIVIRUS.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RA	YIN S.R., PORCELL R.H., EMERSON S.U.;					
RL	SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=K52-87;					
RX	MEDLINE: 95176571.					
RA	YIN S., PORCELL R.H., EMERSON S.U.;					
RT	*A new Chinese isolate of hepatitis E virus: comparison with strains					
RT	recovered from different geographical regions.*;					
RT	VIRUS GENES 9:23-32(1994).					

WPA5RELH (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:19:58 1999; MasPar time 7.22 Seconds
Tabular output not generated. 294.721 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (1-100) from PCTUS9913024.pep (7 of 12)
Perfect Score: 683
Sequence: 1 MEKFMAEFGQGVQTPFLSE.....STSDTTSAAFVSPMNSLTN 100
Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 27.671; Variance 122.380; scale 0.226

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	300	43.9	65	19	R97371	1.46e-17
2	96	14.1	1693	35	W76368	Hepatitis E virus hol
3	92	13.5	520	1	P94617	Neutral protease enco
4	92	13.5	521	3	P51009	Sequence of neutral p
5	89	13.0	438	20	W06355	Malate permease.
6	89	13.0	438	23	W14558	Malate permease.
7	87	12.7	242	19	P95000	Bacillus subtilis lic
8	87	12.7	623	35	W75773	Amino acid sequence o
9	86	12.6	239	1	R06621	Hybrid (1.3-1.4)-pre-
10	86	12.6	1693	37	W81519	Hepatitis E virus (HE
11	86	12.6	1693	36	W80196	Protein encoded by OR
12	86	12.6	1693	10	R51264	HEV strain protein en
13	86	12.6	1693	34	W71209	Protein encoded by OR
14	86	12.6	1693	19	R91813	Hepatitis E virus str
15	85	12.4	633	35	W75774	Amino acid sequence o
16	84	12.3	1693	3	R14618	Protein encoded by OR

17	83	12.2	473	21	W14005	Human SHC protein.	4.33e+01
18	83	12.2	473	15	R84637	SHC protein.	4.33e+01
19	83	12.2	474	17	R97243	SHC phosphotyrosine b	4.33e+01
20	83	12.2	613	15	R74632	OETR ethylene respons	4.33e+01
21	83	12.2	613	36	W73122	A. thaliana ethylene	4.33e+01
22	82	12.0	551	24	W18790	Corrected Bacillus la	5.11e+01
23	82	12.0	633	35	W75775	Amino acid sequence o	5.11e+01
24	81	11.9	345	10	R52824	GTP-cyclohydrolase II	6.03e+01
25	81	11.9	3011	8	R40120	HCV genomic amino aci	6.03e+01
26	81	11.9	3011	12	R66995	Hepatitis C virus gen	6.03e+01
27	80	11.7	170	1	P90150	Sequence of hepatitis	7.11e+01
28	80	11.7	170	1	P92033	Sequence encoded in t	7.11e+01
29	80	11.7	411	16	R90934	HCV NS5 domain antige	7.11e+01
30	80	11.7	499	38	W67010	HCV non-structural pr	7.11e+01
31	80	11.7	509	20	W08378	Brassica napus micros	7.11e+01
32	80	11.7	509	2	R06519	Microspore-specific c	7.11e+01
33	80	11.7	516	7	R33633	HCV CFS-NS5E fusion p	7.11e+01
34	80	11.7	516	4	R21566	HCV CFS-NS5E - pHCV-4	7.11e+01
35	80	11.7	1766	1	P92041	Sequence encoded in t	7.11e+01
36	80	11.7	1786	1	P90158	Protein sequence of h	7.11e+01
37	80	11.7	2261	1	P90164	Peptide encoded by co	7.11e+01
38	80	11.7	2301	1	P92047	Sequence encoded in t	7.11e+01
39	80	11.7	2436	5	R28582	HCV amino acid sequen	7.11e+01
40	80	11.7	2462	1	P90288	Peptide encoded by co	7.11e+01
41	80	11.7	2816	7	R34009	HCV-1 polyprotein.	7.11e+01
42	80	11.7	2955	2	R08124	Hepatitis C virus put	7.11e+01
43	80	11.7	3011	28	W40038	HCV polyprotein.	7.11e+01
44	80	11.7	3011	4	R21519	Compiled HCV sequence	7.11e+01
45	80	11.7	3011	26	W34480	HCV polyprotein.	7.11e+01

ALIGNMENTS

RESULT 1
ID R97371 standard; Protein; 65 AA.
AC R97371;
DT 07-JAN-1997 (first entry)
DE Phase T4 ORFX gene product _gp34.
KW Phase T4; tail fibre protein; nanotechnology; nano-structure;
KW filter; molecular sieve.
OS Bacteriophage T4.
PN W09611947-Al.
PD 25-APR-1996.
PF 13-OCT-1994; US-322760.
PR 13-OCT-1995; U13023.
PA (GOLD/) GOLDBERG E B.
PI Goldberg ES;
DR WPI; 96-221942/22.
DR N-PSDB; T29053.
PT New proteins derived from T4 phage tail fibre proteins - that can self assemble into nano-structure(s), useful as filters etc, also corresponding DNA
PS Claim 7; Fig 7; 83pp; English.
CC A protein (R97371) of unspecified function is the product of open reading frame x of the tail fibre protein gene region (see also T29053) of phage T4. This gene region also includes open reading frames for tail fibre proteins (see also R97370 and R97372-74).
CC Tail fibre proteins (native or modified) can be produced in large quantities in microbial cells and used as building blocks of strong, stable nanostructures.
CC Sequence 65 AA;
Query Match 43.9%; Score 300; DB 19; Length 65;
Best Local Similarity 82.1%; Pred. No. 1.46e-17;
Matches 46; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Db 1 mekfmaefwtricpnallsesnsvrykisiagscplstagspsvkkfdpvgstqtf 56
|||||: :|||||: :|||||: :|||||: :|||||: :|||||: :|||||: :|||||: :|||||: :
QY 1 MEKFMAEFGQGVQTPFLSESNVRYKISAGSCPLSTAGSPVKKFDQPNVGSQTF 56

RESULT 2
ID W76368 standard; Protein; 1693 AA.

```
AC W76368;
DT 03-DEC-1998 (first entry)
DE Hepatitis E virus hollow particle protein #1.
KW Hollow particle protein; virus; antibody; detection; immunoassay;
OS Hepatitis virus.
FH Key Location/Qualifiers
FT Protein 1..1693
FT /note= "Partial sequence"
FT
PN J10234383-A.
PD 08-SEP-1998.
PF 28-FEB-1997; 062445.
PR 28-FEB-1997; JP-062445.
PA (DENK-) DENKA SEIKEN KK.
PA (KOKU-) KOKURITSU YOKO EISEI KENKYUSHO.
DR WPI: 98-535037/46.
DR N-PSDB: V61687.
PT Hepatitis E virus hollow particle poly(peptide(s) and nucleic acids
PT encoding it - useful for more accurate detection of HEV in samples,
PT using immuno-assays and nucleic acid hybridisation
PS Claim 10; Page 17-24; 29pp; Japanese.
CC This sequence represents a Hepatitis E viral hollow particle protein.
CC This polypeptides can be used to raise antibodies to detect HEV
CC infection in samples, e.g. by immuno-assay based techniques, and the
CC nucleic acid can be used for the same in nucleic acid hybridisation
CC assays. The polypeptides and nucleic acids allow more accurate
CC detection of HEV than previously possible.
SQ Sequence 1693 AA;

Query Match 14.1%; Score 96; DB 35; Length 1693;
Best Local Similarity 26.9%; Pred. No. 4.71e+00;
Matches 18; Conservative 21; Mismatches 25; Indels 3; Gaps 3;

Db 572 frtsfvdgavleangperynlfsdqstmaagpfspptyaasaglevryvaagldhrav 631
Qy 8 FQGVYQTPFLSENSVRYKISAGSCPLSTAGPSYVKFQDNPVGSQTP-F-SAGLHLR-V 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 632 fapgvsp 638
| : : :
Qy 65 FDPSTGA 71

RESULT 3
ID P94617 standard; protein; 520 AA.
AC P94617;
DT 21-JUN-1990 (first entry)
DE Neutral protease encoded by npr gene.
KW Protease; expression systems; subtilin; neutral protease; ds.
OS Bacillus amyloliquefaciens.
FH Key Location/Qualifiers
FT Protein 221..520
FT /note="Mature"
FT region 27..520
FT /note="PRO-"
PN US4801537-A.
PD 31-JAN-1989.
PF 29-MAR-1985; 717800.
PR 29-MAR-1985; US-717800.
PA (GENE-) Genex Corp.
PI Nagarajan V, Rhodes CS, Banner CDB;
DR WPI: 89-053639/07.
DR N-PSDB: N91114.
PT Vectors for expression of polypeptide(s) in Bacillus -
PT contg. promoter and regulatory regions which control expression
PT and secretion of protease(s) in Bacillus.
PS Disclosure; p: English.
CC Claimed replicon comprises a promoter and regulatory regions, capable of
CC expressing alkaline and neutral protease genes.
SQ Sequence 520 AA;

Query Match 13.5%; Score 92; DB 1; Length 520;
Best Local Similarity 27.8%; Pred. No. 9.41e+00;
Matches 15; Conservative 19; Mismatches 18; Indels 2; Gaps 2;

AC P94617;
DT 21-JUN-1990 (first entry)
DE Neutral protease encoded by npr gene.
KW Protease; expression systems; subtilin; neutral protease; ds.
OS Bacillus amyloliquefaciens.
FH Key Location/Qualifiers
FT Protein 221..520
FT /note="Mature"
FT region 27..520
FT /note="PRO-"
PN US4801537-A.
PD 31-JAN-1989.
PF 29-MAR-1985; 717800.
PR 29-MAR-1985; US-717800.
PA (GENE-) Genex Corp.
PI Nagarajan V, Rhodes CS, Banner CDB;
DR WPI: 89-053639/07.
DR N-PSDB: N91114.
PT Vectors for expression of polypeptide(s) in Bacillus -
PT contg. promoter and regulatory regions which control expression
PT and secretion of protease(s) in Bacillus.
PS Disclosure; p: English.
CC Claimed replicon comprises a promoter and regulatory regions, capable of
CC expressing alkaline and neutral protease genes.
SQ Sequence 520 AA;

Query Match 13.5%; Score 92; DB 3; Length 521;
Best Local Similarity 27.8%; Pred. No. 9.41e+00;
Matches 15; Conservative 19; Mismatches 18; Indels 2; Gaps 2;

AC P94617;
DT 21-JUN-1990 (first entry)
DE Neutral protease encoded by npr gene.
KW Protease; expression systems; subtilin; neutral protease; ds.
OS Bacillus amyloliquefaciens.
FH Key Location/Qualifiers
FT Protein 221..520
FT /note="Mature"
FT region 27..520
FT /note="PRO-"
PN US4801537-A.
PD 31-JAN-1989.
PF 29-MAR-1985; 717800.
PR 29-MAR-1985; US-717800.
PA (GENE-) Genex Corp.
PI Nagarajan V, Rhodes CS, Banner CDB;
DR WPI: 89-053639/07.
DR N-PSDB: N91114.
PT Vectors for expression of polypeptide(s) in Bacillus -
PT contg. promoter and regulatory regions which control expression
PT and secretion of protease(s) in Bacillus.
PS Disclosure; p: English.
CC Claimed replicon comprises a promoter and regulatory regions, capable of
CC expressing alkaline and neutral protease genes.
SQ Sequence 520 AA;
```

```
Db 239 issesgkyvirdlskptgtqitydlnqreynlp-gtlvsttngftssqraa 291
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 36 LSTAGPSYV-KFQDNPVGSQTFSGAGLHLRVFDPSTGALVDSKSYAFSTSDTTS 88
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
ID P51009 standard; Protein; 521 AA.
AC P51009;
DT 01-DEC-1991 (first entry)
DE Sequence of neutral protease encoded by the npr[Bamp] gene.
KW Bacillus expression vector; secretion vector.
OS Bacillus amyloliquefaciens.
FH Key Location/Qualifiers
FT Peptide 1..27
FT /label= signal
FT peptide 28..221
FT /label= Pro sequence
FT protein 222..521
FT misc_difference 1
FT /label= fMet
PN EP-133756-A.
PD 06-MAR-1985.
PF 06-JUL-1984; 304662.
PR 06-JUL-1983; US-511198.
PR 08-JUN-1984; US-618902.
PR 29-MAR-1985; US-717800.
PA (GENE-) GENEX CORP.
PI Nagarajan V, Banner CDB, Rhodes CS;
DR WPI: 85-057299/10.
DR N-PSDB: N50542.
PT Replicable plasmidic expression vector - for transformation of
PT Bacillus to direct expression of poly:peptide
PS Disclosure; Fig 4; 37pp; English.
CC The inventors claim a vector comprising a replicable plasmid
CC containing the promoter and regulatory region of a gene selected
CC from apr[Bamp] and npr[Bamp], for transformation of Bacillus to
CC direct expression of polypeptide.
SQ Sequence 521 AA;

Query Match 13.5%; Score 92; DB 3; Length 521;
Best Local Similarity 27.8%; Pred. No. 9.41e+00;
Matches 15; Conservative 19; Mismatches 18; Indels 2; Gaps 2;

Db 240 issesgkyvirdlskptgtqitydlnqreynlp-gtlvsttngftssqraa 292
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 36 LSTAGPSYV-KFQDNPVGSQTFSGAGLHLRVFDPSTGALVDSKSYAFSTSDTTS 88
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
ID W06355 standard; Protein; 438 AA.
AC W06355;
DT 13-FEB-1997 (first entry)
DE Malate permease.
KW Malate permease; mael gene; malic acid; succinate; succinic acid;
KW malonate; malonic acid; wine; malolactic fermentation;
KW deacidification; yeast; Saccharomyces cerevisiae.
OS Schizosaccharomyces pombe.
FH Key Location/Qualifiers
FT modified_site 28
FT /label= Phosphorylation
FT /note= "potential protein kinase C
FT phosphorylation site"
FT domain 37..57
FT /label= Transmembrane_domain
FT domain 65..85
FT /label= Transmembrane_domain
FT modified_site 94
FT /label= Phosphorylation
FT /note= "potential protein kinase C
FT phosphorylation site"
FT domain 106..126
FT /label= Transmembrane_domain
```


RESULT 8

ID W75773 standard; Protein: 623 AA.

AC W75773;

DT 02-DEC-1998 (first entry)

DE Amino acid sequence of lepidoteran-active HD573 toxin.

KW HD573 toxin; PCR: primer: amplification; Bacillus thuringiensis; probe: lepidoptera; pest; pesticide; Ostrinia nubilalis; Heliothis virescens; Helicoverpa zea; hybridisation.

KW Bacillus thuringiensis.

OS WO9840490-A1.

PN 17-SEP-1998.

PD 13-MAR-1998: 005081.

PF 13-MAR-1997: US-040512.

PR (MYCO) MYCOGEN CORP.

PA Muller-Cohn J, Narva KE, Schnepf HE;

PI WPI: 98-506734/43.

DR N-PSDB: V52610.

DR New insecticidal Bacillus thuringiensis toxins - useful for controlling lepidopteran pests, especially Ostrinia nubilalis, Heliothis virescens and Helicoverpa zea

PT Claim 14; Pages 28-30; 50pp: English.

PS This is the amino acid sequence of a novel Bacillus thuringiensis toxin used in the method of the invention, to control lepidopteran pests.

CC The new toxins are useful as pesticides, especially for the control of Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The polynucleotide coding sequences are useful for recombinant expression of the toxins and the primers, together with probes derived from the new sequences, are useful for the identification and characterisation of novel genes that encode pesticidal toxins.

CC Sequence 623 AA;

SO

```

Query Match      12.78; Score 87; DB 35; Length 623;
Best Local Similarity 18.23; Pred. No. 2.21e+01;
Matches 12; Conservative 26; Mismatches 26; Indels 2; Gaps 2;

Db 502 fisekyngqdslrfselntartylrgnsgsnlylrvssigsstirvinytany 561
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 12 YVQTPELSESNRYKIS-IAQSCPSTAGPSYVKF-QDNPGVGSQTFSGAHLRVEDPST 69
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 562 ntltnn 567
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 70 GALVDS 75
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT	9	
ID	R06621	standard; protein: 239 AA.
AC	R06621;	
DE	09-JAN-1991	(first entry)
DT	Hybrid (1.3-1.4)	pre-beta-glucanase.
DE	Hybrid (1.3-1.4)	pre-beta-glucanase.
KW	Hybrid pre-beta-glucanase;	glucans; beer; animal feed: poultry.
OS	Bacillus amyloliquefaciens,	Bacillus macerans.
FS	Key	Location/Qualifiers
FT	domain	1..129
FT		/label=amino terminal of beta-amyloliquefaciens
FT	domain	133..236
FT		/label=carboxyl-terminal of B.macerans
PN	WO9009436-A.	
PD	23-AUG-1990.	
PF	16-FEB-1990:	DK0044.
PR	16-FEB-1989:	DD-325800.
PR	04-AUG-1989:	DK-003848.
PA	(CARL-) CARLSBERG A/S.	
PA	(DEAK) AKAD WISSENSCHAFT DDR.	
PI	Borriess R, Hofemeister J, Thomsen KK, Olsen O, Vonwettstein D;	
DR	WPI: 90-275129/36.	
DR	N-PSDB: 005832.	

PT New thermostable (1,3-1,4)-beta-glucanase - prep'd. using hybrid
gene obtd. using *Bacillus amyloliquefaciens* and *B. macerans* genes
disclosure: page 26; 84pp; English.
PT This hybrid protein is encoded by the beta-glucanase-H1 gene.
CC Following processing of the signal peptide the mature protein
CC is produced, comprising the amino terminus of the amylolique-
CC faciens beta-glucanase and the carboxyl-terminal half of the
CC

```

CC B.macerans beta-glucanase. This hybrid protein is thermostable
CC and hydrolyses beta-glycosidic linkages in (1,3-1,4)-beta-glucans.
CC Reducing sugars are obtd. at high temps. and thus this enzyme can
CC be used in the mfr. of food prods., esp. beer and animal feed (eg
CC for feeding poultry). See also Q05833.
SQ Sequence 239 AA;

Query Match 12.6%; Score 86; DB 1; Length 239;
Best Local Similarity 25.5%; Pred. No. 2.62e+01;
Matches 14; Conservative 20; Mismatches 18; Indels 3; Gaps 3;

Db 74 laltspsynkfdcgensrvqtygvlyevrm-kpakntgivyssffitygtptegt 127
   | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 36 LSTAGPSYVKFQ-DNPVGSQTFSAGLH-LRVFDPSTGALVDSKSYAFSTNDTTS 88

RESULT 10
ID W81519 standard; Protein; 1693 AA.
AC W81519;
DT 02-FEB-1999 (first entry)
DE Hepatitis E virus (HEV) ORF-1 protein.
KW Hepatitis E virus; HEV; SAR-55; diagnostic agent; vaccine; antibody;
KW passive immunisation; open reading frame; ORF.

```

```

Query Match      12.68; Score 86; DB 1; Length 239;
Best Local Similarity 25.3%; Pred. No. 2.62e+01;
Matches 14; Conservative 20; Mismatches 18; Indels 3; Gaps 0
Db 74 laitspsynkfcdgenrsqvtgygylgyevrm-kpaknrgivssffitygptgept 127
QV 36 LSTAGSIVYKFFQ-DNPVGSGSTTSAGLH-LRVDFDPTSGALVDSKSYAFSTNDTTS 88

```

RESULT 10
ID W81519 standard; Protein: 1693 AA.

DT 02-FEB-1999 (first entry)
DE Hepatitis E virus (HEV) ORF-1 protein.
KW Hepatitis E virus; HEV; SAR-55; diagnostic agent; vaccine; antibody;
KW passive immunisation; open reading frame; ORF.
OS Hepatitis E virus.

	Key	Misc_difference	Misc_difference	Misc_difference	Misc_difference	Location/Qualifiers
PH	FT	FT	FT	FT	FT	*encoded by /note=1238
PH	FT	FT	FT	FT	FT	*encoded by /note=1244
PH	FT	FT	FT	FT	FT	*encoded by /note=1352
PH	FT	FT	FT	FT	FT	*encoded by /note=1662
PH	FT	FT	FT	FT	FT	*encoded by /note=

Query Match 12.7%; Score 87; DB 35; Length 623;
Best Local Similarity 18.2%; Pred. No. 2.21e+01;
Matches 12; Conservative 26; Mismatches 26; Indels

```

Db 502 fisekyngqgdslrfelsnttarytlrgnqnsynlylrvsigsstirvtngryvtanv 561
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Qy 12 YVQTFPLESSENVBYKIS-IAGSCPILSTAGPSYVKF-QDNPGVSQTFSGAGLHLRVDPST 69

```

Db 562 nttnn 567
QY 70 GALVDS 75

RESULT 9
ID R06621 standard: protein: 239 AA.

CC	protein encoded by a DNA sequence designated SAR-55. SAR-55 also encodes
CC	HEV ORF-2 and ORF-3 proteins. A host organism transformed or transfected
CC	with a recombinant expression vector containing the SAR-55 nucleic acid
CC	can be used to produce the HEV proteins, especially ORF-2 protein. The
CC	recombinant HEV proteins can be used as diagnostic agents and as vaccines
CC	for use against HEV infection. The detection of antibodies specific for
CC	HEV can be used for the diagnosis of infection and diseases caused by
CC	HEV, and for monitoring the progression of such disease. Such methods are
CC	also useful for monitoring the efficacy of therapeutic agents during the
CC	course of treatment of HEV infection and disease in a mammal. The
CC	antibodies can be used for detection or for passive immunisation of
CC	mammals.
SQ	Sequence 1693 AA:

Query Match	12.6%	Score 86;	DB 37;	Length 1693;
Best Local Similarity	25.4%;	Pred. No. 2.62e+01;		
Matches	17; Conservative	22; Mismatches	25; Indels	3; Gaps 3;

```

Db      572 f t s f v d g a v l e t n g p e r h n l s f d a s q s t m a a g p f s l t y a a s a a g l e v r y v a a g l d h r a v   631
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qv      8 F G G G Y V O P F L S E N S V R Y K I S I A G C P L S T A G P S V Y K F O D N P V G S O T - P - S A G I H L R - V   64

```

Db 632 fapgvsp 638

QY 65 FDPSTGA 71

```
RESULT 11
ID W80196 standard; Protein; 1693 AA.
AC W80196;
DE 23-DEC-1998 (first entry)
KW Protein encoded by ORF1 of ET-NANB (HEV) Burma strain DNA sequence.
KW Enterically transmitted non A non B hepatitis virus; ET-NANB;
KW Hepatitis E virus; HEV; Burma HEV isolate; vaccine;
KW diagnostic probe.
OS Non A non B Hepatitis virus.
PN US5824649-A.
PD 20-OCT-1998.
PF 07-JUN-1995; 475807.
PR 25-JUL-1994; US-279823.
PR 17-JUN-1988; US-208997.
PR 11-APR-1989; US-336672.
PR 16-JUN-1989; US-367486.
PR 13-OCT-1989; US-420921.
PR 05-JUL-1990; US-505888.
PR 07-JUN-1995; US-475807.
PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Bradley DW, Fry KE, Krawczynski KZ, Reyes GR, Tam A,
PI Yarbough PO;
DR WPI: 98-582599/49.
DR N-PSDB: V66321.
PT Hepatitis E virus proteins - useful for diagnosis or vaccine
PT production the virus
PS Claim 22; Columns 57-66; 47pp; English.
CC W80196-98 are encoded by the genome of the Burma strain of
CC enterically transmitted non A non B hepatitis virus (ET-NANB)
CC (hepatitis E virus (HEV)). The specification describes an isolated
CC protein which is specifically immunoreactive with antibodies present
CC in individuals infected with HEV and encoded by a sequence contained
CC in an open reading frame (ORF) of an HEV genome. The genome has a
CC sequence that is more than 70% identical to the ORF1 sequence from
CC Burma HEV isolate. The protein is used as a vaccine and a diagnostic
CC probe for ET-NANB.
SQ Sequence 1693 AA;

Query Match 12.6%; Score 86; DB 36; Length 1693;
Best Local Similarity 25.4%; Pred. No. 2.62e+01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 frtsfdvgavletngperhnlfsdagsstnaagpfsityaasaglevryvaagldhrav 631
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 8 FGQGYVQTPLSESNSVRYKISITAGSCPLSTAGPSYVKFQDNVPGSQT-F-SAGLHLR-V 64
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 632 fapgvsp 638
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 65 FDPSTGA 71

RESULT 12
ID R51264 standard; Protein; 1693 AA.
AC R51264;
DE 21-OCT-1994 (first entry)
DE HEV strain protein encoded by ORF-1.
KW Hepatitis E virus; HEV; strain SAR-55; open reading frame; ORF;
KW antibody; detection; diagnosis; primates; stool suspension.
OS Hepatitis E virus strain SAR-55.
PN WO9406913-A.
PD 31-MAR-1994.
PF 17-SEP-1993; U008849.
PR 18-SEP-1992; US-947263.
PA (USSH ) US SEC DEPT HEALTH.
PI Emerson SU, Purcell RH, Tsarev SA;
DR WPI: 94-118462/14.
DR N-PSDB: Q43197.
PT Purified hepatitis E strain SAR-55 virus - used to develop prods.
PT for use in detection, diagnosis, vaccines and therapy of
PT hepatitis E virus infection
PS Disclosure: Page 57-62; 114pp; English.
CC The sequences given in R51264-66 are encoded by the hepatitis E virus
CC (HEV) strain SAR-55. The cDNA sequence contains three open reading
```

```
CC frames (ORFs). These proteins can be used to stimulate the production
CC of protective antibodies upon injection into a mammal that would serve
CC to protect the mammal upon challenge with wild type HEV. The proteins
CC can be used for detection and diagnosis of HEV infection. The HEV
CC SAR-55 cDNA was isolated from primates inoculated with stool
CC suspensions obtained from hepatitis E patients.
SQ Sequence 1693 AA;

Query Match 12.6%; Score 86; DB 10; Length 1693;
Best Local Similarity 25.4%; Pred. No. 2.62e+01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 frtsfdvgavletngperhnlfsdagsstnaagpfsityaasaglevryvaagldhrav 631
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 8 FGQGYVQTPLSESNSVRYKISITAGSCPLSTAGPSYVKFQDNVPGSQT-F-SAGLHLR-V 64
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 632 fapgvsp 638
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 65 FDPSTGA 71

RESULT 13
ID W71209 standard; Protein; 1693 AA.
AC W71209;
DE 30-OCT-1998 (first entry)
DE Protein encoded by ORF 1 of the Burmese isolate of ET-NANB.
KW Enterically transmitted nonA/nonB hepatitis virus; identification;
KW HEV; ET-NANB; detection; vaccine.
OS Hepatitis virus.
FH Key
FT Misc_difference 154 Location/Qualifiers
FT Misc_difference 1514 /note= "not specified"
FT Misc_difference 1514 /note= "not specified"
FT Misc_difference 1552 /note= "not specified"
FT US5789559-A.
PN 04-AUG-1998.
PD 25-JUN-1994; 279823.
PR 05-APR-1991; US-681078.
PR 17-JUN-1986; US-208997.
PR 11-APR-1989; US-336672.
PR 19-JUN-1989; US-367486.
PR 13-OCT-1989; US-420921.
PR 05-APR-1990; US-505888.
PR 25-JUL-1994; US-279823.
PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Bradley DW, Fry KE, Krawczynski KZ, Reyes GR, Tam A,
PI Yarbough PO;
DR WPI: 98-446186/38.
DR N-PSDB: V54729.
PT Hepatitis E virus DNA - useful for e.g. virus detection and viral
PT protein production
PS Disclosure; Columns 55-64; 45pp; English.
CC W71209-11 represent the proteins encoded by the open reading frames
CC (ORFs) of the DNA sequence of the Burmese isolate of an enterically
CC transmitted nonA/nonB viral hepatitis agent (ET-NANB). The nucleic
CC acid sequence may be used for identifying and sequencing the entire
CC viral agent (also referred to as HEV), detecting ET-NANB in
CC infected samples, e.g. by specific amplification of virus-derived DNA
CC sequences and for producing recombinant viral proteins for use in
CC vaccines.
SQ Sequence 1693 AA;

Query Match 12.6%; Score 86; DB 34; Length 1693;
Best Local Similarity 25.4%; Pred. No. 2.62e+01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 frtsfdvgavletngperhnlfsdagsstnaagpfsityaasaglevryvaagldhrav 631
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 8 FGQGYVQTPLSESNSVRYKISITAGSCPLSTAGPSYVKFQDNVPGSQT-F-SAGLHLR-V 64
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 632 fapgvsp 638
```

QY 1 1: :
65 FDPSTGA 71

RESULT 14

ID R91813 standard; Protein: 1693 AA.
AC R91813;
DE 26-NOV-1996 (first entry)
DT Hepatitis E virus strain SAR-55 ORF-1.
KW Hepatitis E virus; HEV; SAR-55 strain; enteric transmission;
KW structural region; antigen; detection; antibody; vaccine;
KW immunisation; infection.
OS Hepatitis E virus.
FH Key
FT Location/Qualifiers
FT misc_difference 1238
FT misc_difference 1455..1693
FT /note= "10 bp nucleic acid sequence TGGTNTTYGA
has to be inserted between nucleotides
4390..4391 of T27394 before these amino
acid residues can be decoded"

PN WO9610580-A2.

PD 11-APR-1996.

PF 03-OCT-1995; U13102.

PR 03-OCT-1994; US-316765.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Emerson SU, Purcell RH, Tsarev SA;

DR WPI: 96-209320/21.

DT N-PSDB; T27394.

PT Isolated and purified hepatitis E virus strain SAR-55 DNA - encodes
PT antigenic protein useful in diagnosis, prophylaxis and treatment of
PT hepatitis E virus infection

PS Disclosure; Pages 9-13; 121pp; English.

CC The present sequence is the protein prod. of ORF-1 from the
CC hepatitis E virus (HEV) strain SAR-55, which was implicated in an
CC enterically transmitted non-A, non-B hepatitis in Pakistan. The
CC protein encoded by the structural region of the virus (i.e. ORF-2),
CC which is capable of forming HEV like particles, is useful for the
CC detection of HEV antibodies (pref. IgG or IgM) in blood, plasma,
CC sera, cerebrospinal fluid, tissue, urine or pleural fluid. The
CC protein, and anti-HEV antibodies generated using the protein, can
CC also be used in vaccines for immunising an animal against HEV
CC infection. The protein is identified as a band of greater than
CC 50 kD following SDS-PAGE of cell lysates of insect cells infected
CC with a HEV ORF-2 contg. baculovirus, i.e. the claimed recombinant
CC expression vectors ppIC9-1779, -1780 and -1781.

SQ Sequence 1693 AA;

Query Match 12.6%; Score 86; DB 19; Length 1693;

Best Local Similarity 25.4%; Pred. No. 2.62e+01;

Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 frtsfdgavletngperhnlstdasgstmaagpslityaasaaglevryvaagldhrav 631

QY 8 FGQGYVOTPFLESNSVRYKISITAGSCPLSTAGPSYVKFQDNVPVGSQT-F-SAGLHLR-Y 64

Db 632 fapgvsp 638

QY 1 1: :
65 FDPSTGA 71

RESULT 15

ID W75774 standard; Protein: 633 AA.

AC W75774;

DE 02-DEC-1998 (first entry)

DT Amino acid sequence of lepidoteran-active HD525 toxin.

KW HD525 toxin; PCR; primer; amplification; Bacillus thuringiensis; probe;

KW Lepidoptera; pest; pesticide; Ostrinia nubilalis; Heliothis virescens;

KW Helicoverpa zea; hybridisation.

OS Bacillus thuringiensis.

PN WO9840490-A1.

PD 17-SEP-1998.

PF 13-MAR-1998; U05081.

PR 13-MAR-1997; US-040512.

PA (MYCO) MYCOGEN CORP.

PI Muller-Cohn J, Narva KE, Schnepf HE;

DR WPI: 98-506734/43.

DR N-PSDB; V52611.

PT New insecticidal Bacillus thuringiensis toxins - useful for

PT controlling lepidopteran pests, especially Ostrinia nubilalis,

PT Heliothis virescens and Helicoverpa zea

PS Claim 14; Pages 32-34; 50pp; English.

CC This is the amino acid sequence of a novel Bacillus thuringiensis toxin
CC used in the method of the invention, to control lepidopteran pests.

CC The new toxins are useful as pesticides, especially for the control of

CC Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The

CC polynucleotide coding sequences are useful for recombinant expression

CC of the toxins and the primers, together with probes derived from the

CC new sequences, are useful for the identification and characterisation

CC of novel genes that encode pesticidal toxins.

SQ Sequence 633 AA;

Query Match 12.4%;

Best Local Similarity 23.7%; Pred. No. 3.10e+01;

Matches 14; Conservative 20; Mismatches 23; Indels 2; Gaps 2;

Db 511 fisekfgnqsdslrfegsnttarytlrgrngnsynlylrvgssignstlrvtgrvvytas 569

QY 12 YVQTPFLSESNVRYKIS-ITAGSCPLSTAGPSYVKF-QDNVPVGSQTSAGLHLRVPDPS 68

Search completed: Wed Sep 1 16:20:34 1999

Job time : 36 secs.


```

Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scofione, F.; Sekiguchi, J.;
Sekowska, A.; Serror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession B69594
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-242 #label KUN
#cross-references GB:299124; GB:AL009126; NID:g2636442; PID:eil84632;
PID:g2636453
#experimental_source strain 168
REFERENCE
A93526
Murphy, N.; McConnell, D.J.; Cantwell, B.A.
Nucleic Acids Res. (1984) 12:5355-5367
The DNA sequence of the gene and genetic control sites for
the excreted B. subtilis enzyme beta-glucanase.
#cross-references MUID:8427222
#accession A22914
#molecule_type DNA
#residues 1-203, 'L', 205-242 #label MUR
#cross-references EMBL:X00754; NID:g39818; PID:g685236
#experimental_source strain C120
#note Lys and CCA for residue 82 as Leu
REFERENCE
A90026
Tezuka, H.; Yuuki, T.; Yabuuchi, S.
Agric. Biol. Chem. (1989) 53:2335-2339
Construction of a beta-glucanase hyperproducing Bacillus
subtilis using the cloned beta-glucanase gene and a
multi-copy plasmid.
#accession A90026
#molecule_type DNA
#residues 1-23, 'S', 25-82, 'S', 84-242 #label TEZ
#cross-references DDBJ:D00518; NID:g216243; PID:g216244
#experimental_source strain Y-25, clone pLE100
REFERENCE
A90027
Yuuki, T.; Tezuka, H.; Yabuuchi, S.
Agric. Biol. Chem. (1989) 53:2341-2346
Purification and some properties of two enzymes from a
beta-glucanase hyperproducing strain, Bacillus subtilis
HL-25
#contents
#note Source was hyperproducing strain HL-25 with gene from strain
Y-25
the amino ends of the mature forms of E-1 and E-2 are
pyroglutamic acid and glutamine, respectively
bglS
catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in
lichenin and other 1,3- and 1,4-beta-linked polysaccharides
#superfamily licheninase
#extracellular protein; glycosidase; hydrolase; polysaccharide
degradation; pyroglutamic acid
, #domain signal sequence #status predicted #label SIG\
#product licheninase #status predicted #label MAT\
#modified site pyrrolidone carboxylic acid (Gln) (in
mature form) (partial) #status experimental
#length 242 #molecular-weight 27268 #checksum 4439
SUMMARY

```

```

Query Match 12.4%; Score 85; DB 1; Length 242;
Best Local Similarity 25.5%; Pred. No. 3.03e+00;
Matches 14; Conservative 20; Mismatches 18; Indels 3; Gaps 3;
Db 77 LATSPAYNKPCCGGRNSQVYGYLYEVRM-KPAKNTGIVSFFTYTGTGCTP 130
|: : : : : | | | | | : : : : : | | | | | : : : : : | | |
QY 36 LSTAGPSYVKFQ-DNPVGSQTFAGLH-LRVFDPSTGALVDKSKYAFSTNSDNTS 88

RESULT 13
ENTRY A61046 #type complete
TITLE ecdysone-induced membrane protein IMP-E3 - fruit fly
(Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
16-Feb-1997
ACCESSIONS A61046
REFERENCE A61046
#authors Moore, J.T.; Fristrom, D.; Hammonds, A.S.; Fristrom, J.W.
#journal Dev. Genet. (1990) 11:299-309
#title Characterization of IMP-E3, a gene active during imaginal
disc morphogenesis in Drosophila melanogaster.
#accession A61046
#status preliminary
#molecule_type mRNA
#residues 1-331 #label MOO
GENETICS FlyBase:ImpE3
#cross-references FlyBase:FBgn0001255
KEYWORDS membrane protein
SUMMARY #length 331 #molecular-weight 36583 #checksum 8221

Query Match 12.4%; Score 85; DB 2; Length 331;
Best Local Similarity 31.9%; Pred. No. 3.03e+00;
Matches 15; Conservative 12; Mismatches 18; Indels 2; Gaps 2;
Db 190 LDFLRLYDDNYGRAFGSESAMDRWSTASIAGKRVRPPTKP-YVDF 235
|: : : : : | | | | | : : : : : | | | | | : : : : : | | |
QY 1 MEKFAEFGQGVQVPFLSESNVRYKI-SIAGSCPLSTAGPSYVKF 46

RESULT 14
ENTRY C56617 #type complete
TITLE cfaC protein precursor - Escherichia coli plasmid NTP113
ORGANISM #formal_name Escherichia coli
DATE 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change
20-Mar-1998
ACCESSIONS C56617
REFERENCE A56617
#authors Jordt, B.J.; Willshaw, G.A.; van der Zeijst, B.A.; Gaastra,
W.
#journal DNA Seq. (1992) 2:257-263
#title The complete nucleotide sequence of region 1 of the CFA/I
fimbrial operon of human enterotoxigenic Escherichia coli.
#cross-references MUID:9232981
#accession C56617
#status preliminary
#molecule_type DNA
#residues 1-869 #label JOR
#cross-references GB:M55661; NID:gl45507; PID:gl45510
#experimental_source enterotoxigenic strain, CFA/I-ST plasmid NTP113
#note sequence extracted from NCBI backbone (NCBIN:108960,
NCBIP:108971)
GENETICS cfaC
#gene plasmid
#genome #length 869 #molecular-weight 97830 #checksum 9755
SUMMARY

Query Match 12.4%; Score 85; DB 2; Length 869;
Best Local Similarity 32.4%; Pred. No. 3.03e+00;
Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;
Db 158 AFIOSQTINLSDSGYKRLSISGNSALGITDTSY 191

```


Result	No.		Score	Query		Length	DB	ID	Description	Pred. No.
	No.	Score		Match						
1	1	300	43.9	65	1	V15A_BP74			HYPOTHETICAL 7.3 KD PR	5.56e-433
2	2	102	14.9	672	1	N003_PARDE			NADH-UBIQUINONE OXIDOR	3.05e-03
3	3	100	14.6	1693	1	POLN_HEYMI			NON-STRUCTURAL POLYPRO	6.34e-03
4	4	92	13.5	521	1	NPRE_BACAM			BACILLOLYSIN PRECURSOR	1.08e-01
5	5	89	13.0	438	1	MAEL_SCHPO			MALIC ACID TRANSPORT P	3.01e-01
6	6	87	12.7	703	1	Y41I_RH1SN			HYPOTHETICAL 76.2 KD P	5.88e-01
7	7	87	12.7	727	1	NUAM_HUMAN			NADH-UBIQUINONE OXIDOR	5.88e-01
8	8	87	12.7	727	1	NUAM_BOVIN			NADH-UBIQUINONE OXIDOR	5.88e-01
9	9	86	12.6	239	1	GUB_BACAM			BETA-GLUCANASE PRECURS	8.18e-01
10	10	86	12.6	457	1	EMB8_PICGL			LATE EMBRYOGENESIS ABU	8.18e-01
11	11	86	12.6	932	1	YMB4_CAEEL			HYPOTHETICAL 104.4 KD	8.18e-01
12	12	86	12.6	1693	1	POLN_HEYBU			NON-STRUCTURAL POLYPRO	8.18e-01
13	13	86	12.6	1693	1	POLN_HEYPA			NON-STRUCTURAL POLYPRO	8.18e-01
14	14	85	12.4	242	1	GUB_BACSU			BETA-GLUCANASE PRECURS	1.14e+00
15	15	85	12.4	869	1	CFAC_ECOLI			CFA/I FIMBRIAL SUBUNIT	1.14e+00
16	16	84	12.3	243	1	GUB_BACLI			BETA-GLUCANASE PRECURS	1.57e+00
17	17	83	12.2	171	1	YAL4_PHVU			HYPOTHETICAL AL4 PROTE	2.17e+00
18	18	83	12.2	578	1	SHC_MOUSE			SHC TRANSFORMING PROTE	2.17e+00
19	19	83	12.2	583	1	SHC_HUMAN			SHC TRANSFORMING PROTE	2.17e+00
20	20	83	12.2	611	1	YX45_MYCTU			HYPOTHETICAL 67.2 KD P	2.17e+00
21	21	83	12.2	960	1	KKIT_CHICK			MAST/STEM CELL GROWTH	2.17e+00
22	22	83	12.2	1325	1	YAE6_SCHPO			HYPOTHETICAL 145.8 KD	2.17e+00
23	23	82	12.0	182	1	Y391_HAEIN			HYPOTHETICAL PROTEIN H	2.99e+00

```

RN  SEQUENCE OF 657-672 FROM N.A.
RX  MEDLINE: 93136200.
RT  "DNA sequencing of the seven remaining structural genes of the gene
RT  cluster encoding the energy-transducing NADH-quinone oxidoreductase
RT  of Paracoccus denitrificans."
RL  BIOCHEMISTRY 32:968-981(1993).
CC  -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
CC  -!- COFACTOR: MAY BIND TWO 4FE-4S CLUSTER AND ONE 2FE-2S CLUSTER.
CC  -!- SUBUNIT: COMPOSED OF 14 DIFFERENT SUBUNITS.
CC  -!- SIMILARITY: BELONGS TO THE COMPLEX I 75 KD SUBUNIT FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: M84572; G150601; -.
DR  PIR: S23948; S23948.
DR  PROSITE: PS00641; COMPLEX1_75K_1; 1.
DR  PROSITE: PS00642; COMPLEX1_75K_2; 1.
DR  PROSITE: PS00643; COMPLEX1_75K_3; 1.
DR  PIR: PF00111; fer2; 1.
DR  PIR: PF00384; molybdopterin; 1.
KW  OXIDOREDUCTASE; NAD; UBIQUINONE; IRON-SULFUR; 4FE-4S.
FT  INIT_MET 0
FT  METAL 25 25 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT  METAL 36 36 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT  METAL 47 47 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT  METAL 50 50 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT  METAL 109 109 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT  METAL 112 112 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT  METAL 118 118 IRON-SULFUR (2FE-2S) (POTENTIAL).
FT  METAL 157 157 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT  METAL 160 160 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT  METAL 163 163 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT  METAL 207 207 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ  SEQUENCE 672 AA; 73028 MW; 40DB761A CRC32;

Query Match 14.9%; Score 102; DB 1; Length 672;
Best Local Similarity 26.0%; Pred. No. 3.05e-03;
Matches 20; Conservative 24; Mismatches 29; Indels 4; Gaps 4;

Db 164 VRFTEVAGITQMGQTGRG-EDSEITSYLNQTLNESMGNIIDLPVGLV-SKPYAFTA 221
QY 24 VRYKISAGSCLSTAGPSYKFDQNPVGSQTFSAAGLHLRVFDP-STGALVDSKSYAFST 82
Db 222 RPWELTKTESIDVMDAL 238
QY 83 SN-DTSAFVSFMSNL 98

RESULT 3
ID POLN HEVNY STANDARD; PRT; 1693 AA.
AC Q04610;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE NON-STRUCTURAL POLYPEPTIDE [CONTAINS: RNA-DIRECTED RNA POLYMERASE
DE (EC 2.7.7.48); HELICASE].
OS HEPATITIS E VIRUS (STRAIN MYANMAR) (HEV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALICIVIRIDAE;
OC CALICIVIRUS.
CC [1]
RN SEQUENCE FROM N.A.
RP MEDLINE: 93227573.
RA AYE T.T., UCHIDA T., MA M.Z., IIDA F., SHIKATA T., ICHIKAWA M.,
RA RIKIHISA T., WINN K.;
RT "Sequence and gene structure of the hepatitis E virus isolated from
```

```

Myanmar."
RT VIRUS GENES 7:95-109(1993).
CC -!- HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF ENTERICALLY
CC TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR  EMBL: D10330; G221705; -.
KW POLYPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; HELICASE;
KW ATP-BINDING.
FT NP_BIND 975 982 ATP (POTENTIAL).
SQ SEQUENCE 1693 AA; 185215 MW; FFCB786D CRC32;

Query Match 14.6%; Score 100; DB 1; Length 1693;
Best Local Similarity 26.9%; Pred. No. 6.34e-03;
Matches 18; Conservative 22; Mismatches 24; Indels 3; Gaps 3;

Db 572 FRTSFVGDVAVLEANGPERYNLSFDASOSTMAAGPFLTYAASAGLEVRYVAAGLDHRAV 631
QY 8 FGQGVVQTFPLSESNVRYKISAGSCLSTAGPSYKFDQNPVGSQI-F-SAGLHLR-V 64
Db 632 FAPGVSP 638
QY 65 FDPSTGA 71

RESULT 4
ID NPBE_BACAM STANDARD; PRT; 521 AA.
AC P06832;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE BACILLOLISIN PRECURSOR (EC 3.4.24.28) (NEUTRAL PROTEASE).
GN NPR.
OS BACILLUS AMYLOLIQUEFACIENS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
CC [1]
RN SEQUENCE FROM N.A.
RP STRAIN-ATCC 23844;
RX MEDLINE: 85006739.
RA VASANTHA N., THOMPSON L.D., RHODES C., BANNER C., NAGLE J.,
RA FILIPULA D.;
RT "Genes for alkaline protease and neutral protease from Bacillus
RT amyloliquefaciens contain a large open reading frame between the
RT regions coding for signal sequence and mature protein."
RL J. BACTERIOL. 159:811-819(1984).
CC -!- FUNCTION: THERMOLABILE EXTRACELLULAR ZINC METALLOPROTEASE.
CC -!- CATALYTIC ACTIVITY: SIMILAR, BUT NOT IDENTICAL, TO THAT OF
CC THERMOLYSIN.
CC -!- COFACTOR: BINDS AND REQUIRES A ZINC ATOM, WHICH IS ESSENTIAL FOR
CC PROTEOLYTIC ACTIVITY. BINDS FOUR CALCIUM IONS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M4 (ZINC METALLOPROTEASE);
CC ALSO KNOWN AS THE THERMOLYSIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR  EMBL: K02497; G143249; -.
DR  PIR: A25415; HYBSN.
DR  PROSITE: PS00142; ZINC_PROTEASE; 1.
```

QY		51	VGSQTFSAGLHLRVDPSTGALVDSKSYAFSTNSDNTTSAFVS-FM	95
RESULT		6		
ID	Y4II_RHISN	STANDARD;	PRT;	703 AA.
AC	P55492;			
DT	01-NOV-1997	(REL. 35,	CREATED)	
DT	01-NOV-1997	(REL. 35,	LAST SEQUENCE UPDATE)	
DT	01-NOV-1997	(REL. 35,	LAST ANNOTATION UPDATE)	
DE	HYPOTHETICAL 76.2 KD PROTEIN Y4II.			
GN	Y4II.			
OS	RHIZOBUM SP. (STRAIN NGR234).			
OG	PLASMID SYM PNGR234A.			
OC	BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;			
OC	RHIZOBIACEAE; RHIZOBIUM.			
[1]				
RN	SEQUENCE FROM N.A.			
RP	MEDLINE; 97305956.			
RX	FREIBERG C.-A., FELLAY R., BAIROCH A., BROUGHTON W.J., ROSENTHAL A.,			
RA	PERRET X.;			
RL	"Molecular basis of symbiosis between Rhizobium and legumes.";			
RT	NATURE 387:394-401(1997).			
CC	- - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).			
CC	- - SIMILARITY: NONE OBVIOUS.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to licens@isb-sib.ch).			
CC				
DR	EMBL: AF000078; G2182446;			
CD	KW HYPOTHETICAL PROTEIN; PLAS MID; TRANSMEMBRANE.			
FT	TRANSEM 23 43			
FT	TRANSEM 69 89			
FT	TRANSEM 143 163			
FT	TRANSEM 250 270			
FT	TRANSEM 357 377			
FT	TRANSEM 432 452			
FT	TRANSEM 644 664			
SQ	SEQUENCE 703 AA; 76183 MW; A2BA53CE CRC32;			
	Query Match 12.7%; Score 87; DB 1: Length 703;			
	Best Local Similarity 26.5%; Pred.No. 5.88e-01;			
	Matches 18; Conservative 20; Mismatches 28; Indels 2; Gaps 2;			
DB	224 SPFDADTSCTSTAATSLFGAAGASTRPPPEWLAFANAALASGAFFIGLSPRVIATATSQA 283			
QY	:: :: : :: : :: : :: : :: : :: : :: : :: : :: : :: : :: :			
Dd	284 KAK-FPIIS 290			
QY	: : : : : 74 DSKSYAFS 81			
RESULT		7		
ID	NUAM_HUMAN	STANDARD;	PRT;	727 AA.
AC	P28331;			
DT	01-DEC-1992	(REL. 24,	CREATED)	
DT	01-DEC-1992	(REL. 24,	LAST SEQUENCE UPDATE)	
DT	01-NOV-1995	(REL. 32,	LAST ANNOTATION UPDATE)	
DE	NADH-UBIQUINONE OXIDOREDUCTASE 75 KD SUBUNIT PRECURSOR (EC 1.6.5.3)			
DE	(EC 1.6.99.3) (COMPLEX I-75KD) (CI-75KD).			
GN	NDUFS1.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUETHERIA;			
OC	PRIMATES; CATARRHINI; HOMINIDAE; HUMO.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 92037608.			
RA	CHOW W., RAGAN J., ROBINSON B.H.;			

```

[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP MEDLINE; 90122814.
RX
RA RUNSCHICK M.J., GENNIS R.B., FEARNLEY I.M., WALKER J.E.;
RT "MITOCHONDRIAL NADH:UBIQUINONE REDUCTASE: COMPLEMENTARY DNA SEQUENCE
OF THE IMPORT PRECURSOR OF THE BOVINE 75-KDA SUBUNIT.";
RL BIOCHEMISTRY 28:9452-9459(1989).
CC -1- FUNCTION: THIS IS THE LARGEST SUBUNIT OF COMPLEX I AND IT IS A
CC COMPONENT OF THE IRON-SULFUR (IP) FRAGMENT OF THE ENZYME. IT
CC MAY FORM PART OF THE ACTIVE SITE DEVICE WHERE NADH IS OXIDIZED.
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE - NAD(+) + UBIQUINOL.
CC -1- COFACTOR: MAY BIND ONE 2FE-2S CLUSTER & ONE 4FE-4S CLUSTER.
CC -1- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 40 DIFFERENT SUBUNITS.
CC -1- SUBCELLULAR LOCATION: MATRIX AND CYTOPLASMIC SIDE OF THE
CC MITOCHONDRIAL INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 75 KD SUBUNIT FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; J02877; G163414; -.
CC PIR; A33552; A33552.
CC DR PROSITE; PS00641; COMPLEX1_75K_1; 1.
CC DR PROSITE; PS00642; COMPLEX1_75K_2; 1.
CC DR PROSITE; PS00643; COMPLEX1_75K_3; 1.
CC DR PFAM; PF00111; fer2; 1.
CC DR PFAM; PF00384; molybdopterin; 1.
CC DR OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION; TRANSIT PEPTIDE;
CC KW IRON-SULFUR; 4FE-4S.
CC FT TRANSIT 1 23 MITOCHONDRION.
CC FT CHAIN 24 727 NADH-UBIQUINONE REDUCTASE 75 KD SUBUNIT.
CC FT METAL 53 53 IRON-SULFUR (4FE-4S) (POTENTIAL) .
CC FT METAL 64 64 IRON-SULFUR (4FE-4S) (POTENTIAL) .
CC FT METAL 75 75 IRON-SULFUR (4FE-4S) (POTENTIAL) .
CC FT METAL 78 78 IRON-SULFUR (4FE-4S) (POTENTIAL) .
CC FT METAL 92 92 IRON-SULFUR (4FE-4S) (POTENTIAL) .
CC FT METAL 128 128 IRON-SULFUR (4FE-4S) (POTENTIAL) .
CC FT METAL 131 131 IRON-SULFUR (4FE-4S) (POTENTIAL) .
CC FT METAL 137 137 IRON-SULFUR (4FE-4S) (POTENTIAL) .
CC FT METAL 176 176 IRON-SULFUR (4FE-4S) (POTENTIAL) .
CC FT METAL 179 179 IRON-SULFUR (4FE-4S) (POTENTIAL) .
CC FT METAL 182 182 IRON-SULFUR (4FE-4S) (POTENTIAL) .
CC FT METAL 226 226 IRON-SULFUR (4FE-4S) (POTENTIAL) .
CC SQ SEQUENCE 727 AA; 79442 MW; 9CAF139C CRC32;

```

```

Best Local Similarity 27.3%  Pred. No. 5.88e-01;
Matches 21; Conservative 18; Mismatches 34; Indels 4; Gaps 4;

Db 183 IRFAS2TAGVDDGTTGCRGN-DMQVGTYIEKMSELSGNIIDICPVGALT-SKPYAFTA 240
QY 24 VRYKISTAGSCPSITAGPSIVKQDNPVGSQTFAGLHLRVDP-STGALVDSKSTAFST 82
Db 241 RPWETRKTESIDVMDAV 257
QY 83 SN-DTTSAAFFVSFMSNL 98

RESULT 9
ID ID GUB_BACAM STANDARD; PRT; 239 AA.
AC P07980;
DT 01-AUG-1988 (REL. 08, CREATED)
DT 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
OS BGLA.
OS BACILLUS AMYLOLIQUEFACIENS.

```


DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE NON-STRUCTURAL POLYPROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE
DE (SC 2.7.7.48); HELICASE].
OS HEPATITIS E VIRUS (STRAIN BURMA) (HEV).
OC VIRUSES; SSNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALICIVIRIDAE;
OC CALICIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92024067.
RA TAM A.W., SMITH M.M., GUERRA M.E., HUANG C.-C., BRADLEY D.W.,
RA FRY K.E., REYES G.R.;
RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT full-length viral genome.";
RL VIROLOGY 185:120-131(1991).
CC -I- HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF ENTERICALLY
CC TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M73218; G330024; -
DR PIR; A40778; MNWHE.
KW ATP-BINDING.
KW POLYPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; HELICASE;
FT NP-BIND 975 982 ATP (POTENTIAL).
SQ SEQUENCE 1693 AA; 185191 MW; C560BE14 CRC32;

Query Match 12.6%; Score 86; DB 1; Length 1693;
Best Local Similarity 25.4%; Pred. No. 8.18e-01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 FRTSFVDGAVLETPGPERHNLSPDASOTMAAGPFSITYAASAGLEVRVVAAGLDHRAV 631
Qy 8 FGQGYVQTPFLESNSVRYKISITAGSCPLSTAGPSYKFDNPGVSGT-F-SAGLHLR-V 64
Db 632 FAPGVSP 638
Qy 65 FDPSTGA 71

RESULT 13
ID POLN-HEVPA STANDARD; PRT; 1693 AA.
AC P33424;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE NON-STRUCTURAL POLYPROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE
DE (SC 2.7.7.48); HELICASE].
OS HEPATITIS E VIRUS (STRAIN PAKISTAN) (HEV).
OC VIRUSES; SSNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALICIVIRIDAE;
OC CALICIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92115700.
RA TSAREV S.A., EMERSON S.U., REYES G.R., TSAREVA T.S., LEGTERS L.J.,
RA MALIK I.A., IQBAL M., PURCELL R.H.;
RT "Characterization of a prototype strain of hepatitis E virus.";
RL PROC. NATL. ACAD. SCI. U.S.A. 89:559-563(1992).
CC -I- HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF ENTERICALLY
CC TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M80581; G329998; -
KW POLYPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; HELICASE;
KW ATP-BINDING.
FT NP-BIND 975 982 ATP (POTENTIAL).
SQ SEQUENCE 1693 AA; 185149 MW; PBCA2483 CRC32;

Query Match 12.6%; Score 86; DB 1; Length 1693;
Best Local Similarity 25.4%; Pred. No. 8.18e-01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 FRTSFVDGAVLETPGPERHNLSPDASOTMAAGPFSITYAASAGLEVRVVAAGLDHRAV 631
Qy 8 FGQGYVQTPFLESNSVRYKISITAGSCPLSTAGPSYKFDNPGVSGT-F-SAGLHLR-V 64
Db 632 FAPGVSP 638
Qy 65 FDPSTGA 71

RESULT 14
ID GUB-BACSU STANDARD; PRT; 242 AA.
AC P04957;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE BETA-GLUCANASE PRECURSOR (EC 3.2.1.73) (ENDO-BETA-1,3-1,4 GLUCANASE)
DE (1,3-1,4-BETA-D-GLUCAN 4-GLUCANOHYDROLASE) (LICHENASE).
GN BGLS OR BGL OR LICS OR N15B.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C120;
RX MEDLINE; 84272222.
RA MURPHY N., MCCONNELL D.J., CANTWELL B.A.;
RT "The DNA sequence of the gene and genetic control sites for the
RT excreted B. subtilis enzyme beta-glucanase.";
RL NUCLEIC ACIDS RES. 12:5355-5367(1984).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HL-25;
RA TEZUKA H., YUUKI T., YABUUCHI S.;
RT "Construction of a beta-glucanase hyperproducing Bacillus subtilis
RT using the cloned beta-glucanase gene and a multi-copy plasmid.";
RL AGRIC. BIOL. CHEM. 53:2335-2339(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-168 / BGSC1A1;
RA YOSHIDA K.-I., SANO H., SEKI S., ODA M., FUJIMURA M., FUJITA Y.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE; 95219081.
RA WOLF M., GECZI A., SIMON O., BORRIS R.;
RT "Genes encoding xylan and beta-glucan hydrolysing enzymes in Bacillus
RT subtilis: characterization, mapping and construction of strains
RT deficient in lichenase, cellulase and xylanase.";
RL MICROBIOLOGY 141:281-290(1995).
RN [5]
RP SEQUENCE OF 1-6 FROM N.A.
RC STRAIN-BR151;
RX MEDLINE; 96178961.
RA SCHNETZ K., STUELKE J., GERTZ S., KRUGER S., KRIEG M., HECKER M.,
RA RAK B.;
RT "LicT, a Bacillus subtilis transcriptional antiterminator protein of
RT the BglG family.";
RL J. BACTERIOL. 178:1971-1979(1996).
RN [6]

[illegible]

```

RT *The nucleotide sequence of the first two genes of the CFA/I fimbrial
RT operon of human enterotoxigenic Escherichia coli." ;
RL MICROS. PATHOG. 6:297-309(1989).
RP [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92329981.
RA JORDI B.J.A.M., WILLSHAW G.A., VAN DER ZEIJST B.A.M., GAASTRA W.;
RA "The complete nucleotide sequence of region 1 of the CFA/I fimbrial
RT operon of human enterotoxigenic Escherichia coli." ;
RL DNA SEQ. 2:257-263(1992).
CC -1- FUNCTION: MAY SERVE AS ANCHOR FOR THE FIMBRIAE IN THE OUTER
CC MEMBRANE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M55661; GI45510; -.
KW ANTIGEN; SIGNAL; FIMBRIA; OUTER MEMBRANE; PLASMID.
FT SIGNAL 1 22 POTENTIAL
FT CHAIN 23 869 CFA/I FIMBRIAL SUBUNIT C.
SQ SEQUENCE 869 AA; 97830 MW; 7AF76347 CRC32;
Query Match 12.4%; Score 85; DB 1; Length 869;
Best Local Similarity 32.4%; Pred. No. 1.14e+00;
Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;

Db 158 AFISQOTINLSDSGKYRRLSIGNSALGITDTSY 191
::: : : : : : : : : : : : : : : : : :
QY 11 GYVQTPLESNSVRYK-ISIAGSCPLSTAGPSY 43

Search completed: Wed Sep 1 16:18:21 1999
Job time : 13 secs.

```

WIREH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Sep 1 16:18:39 1999; MasPar time 9.82 Seconds
Tabular output not generated. 556.004 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (1-100) from PCTUS9913024.pep (7 of 12)
Perfect Score: 683
Sequence: 1 MEKFMAEFGQGVQTPFLSE.....STSDNTTSAAPVSMNSLTN 100

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sprenb19

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 38.557; Variance 74.296; scale 0.519

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	101	14.8	1577	2	054178 GLUCOSYLTRANSFERASE.	1.82e-02
2	100	14.6	1593	14	081876 COMPLETE GENOME SEQUEN	2.59e-02
3	96	14.1	114	14	041088 A506L PROTEIN.	9.70e-02
4	92	13.5	229	10	P93830 EARLY AUXIN-INDUCED PR	3.53e-01
5	92	13.5	299	14	065574 HYPOTHETICAL 32.4 KD P	3.53e-01
6	92	13.5	300	14	065581 UL7 POLYPEPTIDE.	3.53e-01
7	92	13.5	337	2	044647 NEUTRAL PROTEASE (FRAG	3.53e-01
8	91	13.3	229	10	049162 IAA17/AXR3-1 PROTEIN.	4.85e-01
9	90	13.2	3443	14	011979 POLYPROTEIN.	6.65e-01
10	89	13.0	566	2	084354 HYPOTHETICAL 63.5 KD P	9.10e-01
11	89	13.0	921	5	P90770 C34B7.2 PROTEIN.	9.10e-01
12	88	12.9	180	10	049984 PUTATIVE ETHYLENE REGE	1.24e+00
13	88	12.9	1546	2	045996 SCAFFOLDING PROTEIN PR	1.24e+00
14	87	12.7	239	2	045691 ENDO-BETA-1,3-1,4-GLUC	1.69e+00
15	86	12.6	521	2	044677 NEUTRAL PROTEASE.	2.30e+00
16	86	12.6	574	5	Q18470 C35A5.2 PROTEIN.	2.30e+00
17	86	12.6	974	5	Q23510 F54G8.4 PROTEIN.	2.30e+00
18	86	12.6	1693	14	069410 METHYL TRANSFERASE.	2.30e+00
19	86	12.6	1893	14	Q81344 HEPATITIS E VIRUS COM	2.30e+00
20	86	12.6	1693	14	Q89444 UNNAMED PROTEIN PRODUC	2.30e+00

21	86	12.6	1693	14	081862 ORF 1.	2.30e+00
22	84	12.3	622	2	Q45743 CRYIIC DELTA-ENDOTOXIN	4.22e+00
23	83	12.2	262	2	P96404 HYPOTHETICAL 27.3 KD P	5.69e+00
24	83	12.2	281	9	Q38022 ORF 1.	5.69e+00
25	83	12.2	461	3	P79065 NOCI PROTEIN.	5.69e+00
26	83	12.2	554	5	Q23267 PROBABLE CARBOXYLESTER	5.69e+00
27	83	12.2	583	4	O15290 P66SHC.	5.69e+00
28	83	12.2	613	10	Q38846 ETHYLENE RESPONSE SENS	5.69e+00
29	83	12.2	673	5	Q01303 ZC375.3 PROTEIN.	5.69e+00
30	83	12.2	1693	14	Q39221 NONSTRUCTURAL POLYPROT	5.69e+00
31	82	12.0	396	2	Q68892 METALLOPROTEASE TOXIN-	7.66e+00
32	82	12.0	486	2	P94412 HOMOLOGUE OF HYPOTHETI	7.66e+00
33	82	12.0	744	3	Q01388 NADH DEHYDROGENASE SUB	7.66e+00
34	82	12.0	811	14	Q36185 CAPSID PROTEIN PRECURS	7.66e+00
35	81	11.9	196	14	Q65847 PUTATIVE COAT PROTEIN.	1.03e+01
36	81	11.9	216	2	Q87609 FLAGELLAR L-RING PROTE	1.03e+01
37	81	11.9	377	2	Q83762 HYPOTHETICAL 41.7 KD P	1.03e+01
38	81	11.9	396	2	Q68424 PUTATIVE METALLOPROTEA	1.03e+01
39	81	11.9	433	2	Q31312 ENOLASE (EC 4.2.1.11)	1.03e+01
40	81	11.9	613	10	Q49231 ETHYLENE RESPONSE SENS	1.03e+01
41	81	11.9	633	2	Q52764 INSECTICIDAL CRYSTAL P	1.03e+01
42	81	11.9	842	14	Q65848 PUTATIVE READTHROUGH P	1.03e+01
43	81	11.9	803	10	Q23540 PROTEIN KINASE HOMOLOG	1.03e+01
44	81	11.9	3011	14	Q03463 GENOME POLYPROTEIN.	1.03e+01
45	80	11.7	275	14	Q68469 POLYPROTEIN (FRAGMENT)	1.38e+01

ALIGNMENTS

RESULT	1
ID	Q54178 PRELIMINARY; PRT; 1577 AA.
AC	Q54178; Q54247;
DT	01-NOV-1996 (TREMREL. 01, CREATED)
DT	01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT	01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE	GLUCOSYLTRANSFERASE.
GN	GTFC.
OS	STREPTOCOCCUS GORDONII.
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC	STREPTOCOCCUS.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-CHALLIS;
RX	MEDLINE: 96157084.
RA	VICKERMAN M.M., SULAVIK M.C., CLEWELL D.B.;
RT	*Molecular analysis of Streptococcus gordonii glucosyltransferase phase variants.*
RL	DEV. BIOL. STAND. 85:309-314(1995).
RN	[2]
RP	SEQUENCE OF 1-96 FROM N.A.
RC	STRAIN-CHALLIS;
RX	MEDLINE: 92276337.
RA	SULAVIK M.C., TARDIF G., CLEWELL D.B.;
RT	*Identification of a gene, rgg, which regulates expression of glucosyltransferase and influences the Spp phenotype of Streptococcus gordonii Challis.*
RL	J. BACTERIOL. 174:3577-3586(1992).
DR	EMBL: U12843; GI054877; -.
DR	EMBL: M89776; GI53795; -.
DR	PFAM: PF00128; alpha-amylase; 1.
KW	TRANSFERASE.
SQ	SEQUENCE 1577 AA; 177805 MW; 7D58FEC2 CRC32;
Query Match	14.8%; Score 101; DB 2; Length 1577;
Best Local Similarity	34.0%; Pred. No. 1.85e-02;
Matches	16; Conservative 10; Mismatches 19; Indels 2; Gaps 1;
Db	239 ODDGTGVKNFAVELNGKILYFDAETGALVDSNEYOFQGTSSLNNEF 285
QY	47 QDNVGSQTFAGLHLRV--FDPSTGALVDSKSYAFSTNDTTSAAF 91
RESULT	2

```
ID Q81876 PRELIMINARY; PRT; 1693 AA.
AC Q81876;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE COMPLETE GENOME SEQUENCE.
OS HEPATITIS E VIRUS (HEV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALICIVIRIDAE;
OC CALICIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UIGH179;
RA UCHIDA I.;
RL SUBMITTED (MAY-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: D11093; D1002342; -
SQ SEQUENCE 1693 AA; 184840 MW; 8952DD38 CRC32;

Query Match 14.68; Score 100; DB 14; Length 1693;
Best Local Similarity 26.98; Pred. No. 2.59e-02;
Matches 18; Conservative 22; Mismatches 24; Indels 3; Gaps 3;

Db 572 FRTSFVGVAGLEANGPERYNLSFASQSTMAAGPFLTYAAGLEVRVVAAGLDHRAV 631
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 8 FGQGYVOTPELSESNVRYKISAGSCPLSTAGPSYVKFQDNPVGSQT-P-SAGLHUR-V 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 632 PAPCVSP 638
| : : :
QY 65 FDPSTGA 71

RESULT 3
ID Q41088 PRELIMINARY; PRT; 114 AA.
AC Q41088;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE A606L PROTEIN.
DE A606L.
GN A606L.
OC PARAMECIUM BURSARIA CHLORELLA VIRUS 1 (PBCV-1).
OS VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PHYCODNAVIRIDAE; PHYCODNAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98022962.
RA LI Y., LU Z., SUN L., ROPP S., KUTISH G.F., ROCK D.L.,
VAN ET TEN J.L.;
RT "Analysis of 74 kb of DNA located at the right end of the 330-kb
chloroella virus PBCV-1 genome.";
RL VIROLOGY 237:360-377(1997).
DR EMBL: U42580; G2447164; -
SQ SEQUENCE 114 AA; 13257 MW; 1B75BE43 CRC32;

Query Match 14.18; Score 96; DB 14; Length 114;
Best Local Similarity 27.48; Pred. No. 9.70e-02;
Matches 17; Conservative 18; Mismatches 25; Indels 2; Gaps 1;

Db 5 FSSMSSNISFRVRLTASCAT--LNPSVGRGRKSRNNLSNFTMSLMLVFVISKPSMSIK 62
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 17 FLSNSVRYKISAGSCPLSTAGPSYVKFQDNPVGSQTFAGLHLRVFDPSTGALVDSK 76
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 KY 64
QY 77 SY 78

RESULT 4
ID P93830 PRELIMINARY; PRT; 229 AA.
AC P93830;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DE EARLY AUXIN-INDUCED PROTEIN IAA17.
DE EARLY AUXIN-INDUCED PROTEIN IAA17.
GN IAA17 OR F19P19.31 OR AXR3.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
```

```
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98004476.
RA KIM J., HARTER K., THEOLOGIS A.;
RT "Protein-protein interactions among the Aux/IAA proteins.";
RL PROC. NATL. ACAD. SCI. U.S.A. 94:11786-11791(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA VYOTSKAIA V.S., OSBORNE B.I., TORIUMI M., YU G., QUI O., BUEHLER E.,
RA CONWAY A.B., CONWAY A.R., DEWAR K., FENG J., KIM C., KURTZ D., LI Y.,
RA SHINN P., SUN H., DAVIS R.W., ECKER J.R., FEDERSPIEL N.A.,
RA THEOLOGIS A.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RX MEDLINE: 98146427.
RA ROUSE D., MACKAY P., STIRNBERG P., ESTELLE M., LEYSER O.;
RT "Changes in auxin response from mutations in an AUX/IAA gene.";
RL SCIENCE 279:1371-1373(1998).
DR EMBL: U49073; G2618723; -
DR EMBL: AC000104; G2341044; -
DR EMBL: AF040631; G2921756; -
DR MENDEL: 7119; ARACH:1524; 6.
FT CONFLICT 2 2 MISSING (IN REF. 2).
FT CONFLICT 197 197 P -> PFIVSVKIRALFLFI (IN REF. 2).
FT CONFLICT 218 218 A -> GMVEL (IN REF. 2).
FT CONFLICT 220 229 RAMEKCKSRA -> MHIIFTLIKF (IN REF. 2).
SQ SEQUENCE 229 AA; 25288 MW; 00A134E6 CRC32;

Query Match 13.58; Score 92; DB 10; Length 229;
Best Local Similarity 21.88; Pred. No. 3.53e-01;
Matches 17; Conservative 23; Mismatches 33; Indels 5; Gaps 5;

Db 88 PVSRYKNVMVSCOKSGGPEAAAFVKVNSMDGAPYLKRLDRLMYK-SYDELSNLSNMF 146
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 23 SVR-YKISTAGSCPLSTAGPSYVKFQDNPVGSQTFAGLHLRVFDPSTGALVDS-KS-YA 79
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 147 SFTMGKHGGEGEMIDFMN 164
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 80 -FSTSDNTTSAAFVSFMN 96
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
ID Q65574 PRELIMINARY; PRT; 299 AA.
AC Q65574;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 32.4 KD PROTEIN.
GN UL7.
OS BOVINE HERPESVIRUS TYPE 1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COOPER, AND JURA;
RX MEDLINE: 95313343.
RA VLCEK C., BENES V., LU Z., KUTISH G.F., PACES V., ROCK D.,
RA LETCHWORTH G.J., SCHWYZER M.;
RT "Nucleotide sequence analysis of a 30-kb region of the bovine
herpesvirus 1 genome which exhibits a colinear gene arrangement with
the UL21 to UL4 genes of herpes simplex virus.";
RL VIROLOGY 210:100-108(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-JURA;
RA SCHWYZER M., PACES V., LETCHWORTH G.J., MISRA V., BUHK H.J.,
RA LOWERY D.E., SIMARD C., BELLO L.J., THIRY E., VLCEK C.;
```

RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-JURA;
RA SCHWYZER M.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; 248053; G971327; -;
KW EMBL; AJ004801; E1187337; -;
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 299 AA; 32379 MW; A3897D4A CRC32;

Query Match 13.5%; Score 92; DB 14; Length 299;
Best Local Similarity 47.2%; Pred. No. 3.53e-01;
Matches 17; Conservative 9; Mismatches 6; Indels 4; Gaps 4;

Db 45 PRFVCEVREIPAGPTFTSSSITHLRV-EPSTGALL 79

Qy 41 PSYV-KFQDNVGSQTF-SAGL-HLRVDFPSTGALV 73

RESULT 6
ID Q65581 PRELIMINARY; PRT; 300 AA.
AC Q65581;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE UL7 POLYPEPTIDE.

GN UL7
OS BOVINE HERPESVIRUS TYPE 1.
OC VIRUSES: DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-SCHOENBOEREN;

RX MEDLINE; 96135223

RA SCHMITT J., KEIL G.M.;

RT "Identification and characterization of the bovine herpesvirus 1 UL7 gene and gene product which are not essential for virus replication in cell culture.";

RL J. VIROL. 70:1091-1099(1996).

DR EMBL; X91751; G1006630; -;

SQ SEQUENCE 300 AA; 32450 MW; 86377347 CRC32;

Query Match 13.5%; Score 92; DB 14; Length 300;
Best Local Similarity 47.2%; Pred. No. 3.53e-01;
Matches 17; Conservative 9; Mismatches 6; Indels 4; Gaps 4;

Db 46 PRFVCEVREIPAGPTFTSSSITHLRV-EPSTGALL 80

Qy 41 PSYV-KFQDNVGSQTF-SAGL-HLRVDFPSTGALV 73

RESULT 7
ID Q44647 PRELIMINARY; PRT; 337 AA.
AC Q44647;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE NEUTRAL PROTEASE (FRAGMENT).

OS BACILLUS AMYLOLIQUEFACIENS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.

RN [1]

RP SEQUENCE FROM N.A.

RA YOSHIMURA K., MIYAZAKI T., NAKAHAMA K., KIKUCHI M.;

RL TAKEDA KENYUSHO HO 44:42-50(1985).

DR EMBL; M64815; G142890; -;

KW PROTEASE.

FT NON_TER 337

SQ SEQUENCE 337 AA; 36910 MW; 5A9BC52B CRC32;

Query Match 13.5%; Score 92; DB 2; Length 337;
Best Local Similarity 27.8%; Pred. No. 3.53e-01;

Matches 15; Conservative 19; Mismatches 18; Indels 2; Gaps 2;

Db 240 ISSESGKVVLRDLSKPTGQIITYDLQREYNLP-GLTVSSSTNQPTTSSQRAA 292

Qy 36 LSTAGPSIV-KFQDNVGSQTF-SAGLHLRVDFPSTGALVDSKSYAFSTSNQITS 88

RESULT 8

ID O49162 PRELIMINARY; PRT; 229 AA.

AC O49162;

DT 01-JUN-1998 (TREMBLREL. 06, CREATED)

DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE IAA17/AXR3-1 PROTEIN.

GN AXR3.

OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).

OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;

OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;

OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CV. COLUMBIA;

RX MEDLINE; 98146427.

RA ROUSE D., MACKAY P., STIRNBERG P., ESTELLE M., LEYSER O.;

RT "Changes in auxin response from mutations in an AUX/IAA gene.";

RL SCIENCE 279:1371-1373(1998).

DR EMBL; AF040632; G2921758; -;

SQ SEQUENCE 229 AA; 25304 MW; 2D88B037 CRC32;

Query Match 13.3%; Score 91; DB 10; Length 229;

Best Local Similarity 22.1%; Pred. No. 4.85e-01;

Matches 17; Conservative 22; Mismatches 33; Indels 5; Gaps 5;

Db 89 VRSYRKNWVSCQKSGGPEAAAFVKVSDGAPYLKIDLRMYK-SYDELSNALSNNFSS 147

Qy 24 VR-YKISAGSCLPSTAGPSYVKFQDNVGSQTF-SAGLHLRVDFPSTGALVDS-KS-YA- 79

Db 148 FTMGRKHGEGEIMDFWN 164

Qy 80 FSTSDNTTSAAFVSPFN 96

RESULT 9

ID O11979 PRELIMINARY; PRT; 3443 AA.

AC O11979;

DT 01-JUL-1997 (TREMBLREL. 04, CREATED)

DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE POLYPROTEIN.

OS MAIZE CHLOROTIC DWARF WAIKAVIRUS.

OC VIRUSES: SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; SEQUIVIRIDAE;

OC WAIKAVIRUS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-TENNESSEE (TN);

RX REDDICK B.B., HABERA L.F., LAW M.D.;

RT "Nucleotide sequence and taxonomy of maize chlorotic dwarf virus

within the family Sequiviridae.";

RL J. GEN. VIROL. 78:1165-1174(1997).

DR EMBL; U67839; G2160664; -;

DR PFAM; PF00680; RNA_dep_RNA_pol; 1.

DR PFAM; PF00910; RNA_helicase; 1.

KW POLYPROTEIN.

SQ SEQUENCE 3443 AA; 387659 MW; 7E60C323 CRC32;

Query Match 13.2%; Score 90; DB 14; Length 3443;

Best Local Similarity 31.8%; Pred. No. 6.65e-01;

Matches 14; Conservative 11; Mismatches 19; Indels 0; Gaps 0;

Db 1193 VDSQERRYGYLRIWHPDNGSLDEGVFAMSTNLEPPFGAFVKY 1236

Qy 51 VGSQTF-SAGLHLRVDFPSTGALVDSKSYAFSTSDNTTSAAFVSF 94

[illegible]

Qy 8 FGQGYVQTPF 17

RESULT 2

ID R40095 standard; peptide; 43 AA.
 AC R40095; 1994 (first entry)
 DE Hib OMP P1-P2 hybrid peptide lp13-2P2.
 DE Haemophilus influenzae; type b; Hib; outer membrane protein; P1; P2;
 KW P6; vaccine; antibody; detection; lipoglycopeptide conjugate;
 KW immunogen.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT region 1..14
 FT /label= C-P1
 FT region 15..43
 FT /label= CHIBP2
 PN W09315205-A.
 PD 05-AUG-1993. CA0041.
 PF 03-FEB-1993; GB-002219.
 PR 03-FEB-1992; GB-002219.
 PA (CONN-) CONNAUGHT LAB LTD.
 PI Chong P, Kandil A, Klein MH, Sia C;
 PI WPI: 93-258681/32.
 DR Synthetic Haemophilus influenzae conjugate vaccine - comprising
 PT T-helper cell determinants and B-cell epitope(s) linked to
 PT synthetic oligo:saccharide(s)
 PS Table 11; Page 59; 99pp; English.
 CC The sequences given in R40053-101 are peptide fragments derived from
 CC the Haemophilus influenzae type b (Hib) outer membrane proteins P1,
 CC P2 and P6. These peptides may be used in a vaccine against Hib
 CC infection and antibodies against these peptides may be used in test
 CC kits to detect H. influenzae in a sample. The vaccine may further
 CC comprise an immunogenic or immunostimulatory molecule or the peptides
 CC may be modified with lipids, or linked to synthetic PRP as synthetic
 CC lipoglycopeptide conjugates to produce alternative vaccines.
 SQ Sequence 43 AA.

Query Match 64.6%; Score 53; DB 8; Length 43;
 Best Local Similarity 50.0%; Pred. No. 7.02e+01;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 20 yaqgyletrf 29

Qy 8 FGQGYVQTPF 17

RESULT 3

ID R07043 standard; protein; 341 AA.
 AC R07043; 1990 (first entry)
 DE P3 gene product of Haemophilus strain 8358 (OMP subtype 6U).
 DE P3 gene; strain 1H; influenza; vaccine; ds.
 OS Haemophilus influenzae.
 PN EP-378929-A.
 PD 25-JUL-1990.
 PF 22-DEC-1989; 313573.
 PR 23-DEC-1988; GB-030124.
 PR 01-FEB-1989; GB-002178.
 PA (CONN-) CONNAUGHT LAB LTD.
 PI Munson RS, Tolan RW, Chong P, Fahim R, McVerry P, Klein M;
 PI WPI: 90-225607/30.
 DR N-PSDB; Q06120.
 PT Gene coding for protein P2 of Haemophilus influenzae type-B -
 PT used for developing vaccines for protection against disease
 PT caused by the organism.
 PS Claim 1; Page 8; 15pp; English.
 CC Peptide may be used as a vaccine to the disease caused by H1
 CC type b, as carrier for conjugation to oligosaccharide derived
 CC from Haemophilus. Protein may be used with other haptens as
 CC T-cell dependant antigen and carrier.
 SQ Sequence 341 AA;

Db 58 yaqgyletrf 67

Qy 8 FGQGYVQTPF 17

RESULT 4

ID R07042 standard; protein; 341 AA.
 AC R07042;
 DT 04-DEC-1990 (first entry)
 DE P3 gene product of Haemophilus strain Durot (OMP subtype 2L).
 DE P3 gene; strain 1H; influenza; vaccine; ds.
 OS Haemophilus influenzae.
 PN EP-378929-A.
 PD 25-JUL-1990.
 PF 22-DEC-1989; 313573.
 PR 23-DEC-1988; GB-030124.
 PR 01-FEB-1989; GB-002178.
 PA (CONN-) CONNAUGHT LAB LTD.
 PI Munson RS, Tolan RW, Chong P, Fahim R, McVerry P, Klein M;
 PI WPI: 90-225607/30.
 DR N-PSDB; Q06119.
 PT Gene coding for protein P2 of Haemophilus influenzae type-B -
 PT used for developing vaccines for protection against disease
 PT caused by the organism.
 PS Claim 1; Page 8; 15pp; English.
 CC Peptide may be used as a vaccine to the disease caused by H1
 CC type b, as carrier for conjugation to oligosaccharide derived
 CC from Haemophilus. Protein may be used with other haptens as
 CC T-cell dependant antigen and carrier.
 SQ Sequence 341 AA;

Query Match 64.6%; Score 53; DB 1; Length 341;
 Best Local Similarity 50.0%; Pred. No. 7.02e+01;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 58 yaqgyletrf 67

Qy 8 FGQGYVQTPF 17

RESULT 5

ID R66880 standard; Protein; 342 AA.
 AC R66880;
 DT 22-AUG-1995 (first entry)
 DE H. influenzae protein P2 encoded by pNV-6.
 DE Haemophilus influenzae type b; outer membrane protein; P2; pNV-6.
 OS Haemophilus influenzae type b.
 PN W09503069-A.
 PD 02-FEB-1995.
 PF 22-JUL-1994; U08326.
 PR 23-JUL-1993; US-096181.
 PA (NAVA-) NORTH AMERICAN VACCINE INC.
 PI Liang S, Pullen JK, Soper TS, Tai JV;
 PI WPI: 95-075024/10.
 DR N-PSDB; Q84357.
 PT Production of Haemophilus influenzae HibP2 protein - for the
 PT prevention of bacterial meningitis in animals
 PS Claim 15; Figure 7; 63pp; English.
 CC To clone P2 into expression vector pET-11a (Novagen pET System
 CC Manual) a 1.1kb fragment from full length P2 (in pNV-1) (see
 CC Q84354) was generated using PCR and oligos Q85350 and Q84353.
 CC Oligo Q85350 allowed the mature P2 protein to be cloned into the NdeI
 CC site of pET 11a. Q84353 incorporated a BamHI site. The P2 fragment
 CC thus generated was digested with NdeI and BamHI, purified and
 CC ligated into pET11a previously cut with NdeI-BamHI. This resulted
 CC in a mature P2 construct - pNV-6. The rest of the plasmid is
 CC identical to pET 11a.
 SQ Sequence 342 AA;

```

Query Match          64.6%; Score 53; DB 13; Length 342;
Best Local Similarity 50.0%; Pred. No. 7.02e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 59 yaqgyletrf 68
   :|||:|
QY 8 FGQGYVQTFF 17

RESULT
ID R66877 standard; Protein; 342 AA.
AC R66877;
DE 22-AUG-1995 (first entry)
DE H. influenzae protein P2 encoded by pNV-3.
KW Haemophilus influenzae type b; outer membrane protein; P2; pNV-3.
OS Haemophilus influenzae type b.
PN W09503069-A.
PD 02-FEB-1995.
PF 22-JUL-1994; U08326.
PR 23-JUL-1993; US-096181.
PA (NAVA-) NORTH AMERICAN VACCINE INC.
PI Liang S, Pullen JK, Soper TS, Tai JY;
DR WPI: 95-075024/10.
DR N-PSDB: Q84356.
PT Production of Haemophilus influenzae HibP2 protein - for the
PT prevention of bacterial meningitis in animals
PS Claim 13; Figure 6; 63pp; English.
CC pNV-3 contains mature P2 contg. a Met and the N-terminus.
CC To clone P2 into expression vector pET-17b (Novagen PET System
CC Manual) the original P2 gene (in pNV-1) (see Q84354) was modified
CC that allowed the mature protein to be cloned into the NdeI site
CC of pET 17b, thus producing the mature P2. The oligo designed for
CC this is Q84350. The extraneous 3' sequences were eliminated by
CC introducing xhoI site about 40 bp from the translation
CC stop codon. This oligo (Q84352) contd. an xhoI site to allow it to
CC be cloned into the xhoI site of pET-17b. PCR was used to generate a
CC 1.1 kb fragment from the full length P2 (pNV-1) with the 5' oligo
CC contg. the NdeI site and the 3' oligo contg. the xhoI site.
CC This fragment was digested with NdeI and xhoI, purified and ligated
CC into the NdeI-xhoI digested pET-17b. This yielded the mature P2
CC construct pNV-3.
SQ Sequence 342 AA;

Query Match          64.6%; Score 53; DB 13; Length 342;
Best Local Similarity 50.0%; Pred. No. 7.02e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 59 yaqgyletrf 68
   :|||:|
QY 8 FGQGYVQTFF 17

RESULT
ID R66877 standard; Protein; 361 AA.
AC R66877;
DE 22-AUG-1995 (first entry)
DE H. influenzae P2 encoded by pNV-1 cDNA.
KW Haemophilus influenzae type b; outer membrane protein; P2.
OS Haemophilus influenzae type b.
PN W09503069-A.
PD 02-FEB-1995.
PF 22-JUL-1994; U08326.
PR 23-JUL-1993; US-096181.
PA (NAVA-) NORTH AMERICAN VACCINE INC.
PI Liang S, Pullen JK, Soper TS, Tai JY;
DR WPI: 95-075024/10.
DR N-PSDB: Q84354.
PT Production of Haemophilus influenzae HibP2 protein - for the
PT prevention of bacterial meningitis in animals
PS Example; Figure 4; 63pp; English.
CC Total genomic DNA was isolated from H. influenzae type b strain Eagan
CC and used as a template for two P2 specific oligos in a PCR. The 5'

```

```

oligo (Q84348) was designed to be 40 bp 5' of the ATG start codon.
The 3' oligo (Q84349) was designed to be 300 bp 3' of the stop codon.
Both oligos contd. SalI restriction sites. The amplified DNA was
used to transform E.coli and the resulting colonies isolated and
analysed. Clones contg. a 1.4 kb fragment were chosen for DNA
sequence analysis. One clone, designated pNV-1, was found to be
identical to the published sequence for Hib strain Minn A (Munson,
R. and Tolan, R.W, Infection & Immunity 57:88-94, Jan. 1989). The
synthetic oligos used to sequence the DNA are indicated in the FT.
The direction of sequencing was 5'-3'. The rest of pNV-1 is
identical to pUC18. The lac promotor is adjacent to the lower
SalI site.
SQ Sequence 361 AA;

Query Match          64.6%; Score 53; DB 13; Length 361;
Best Local Similarity 50.0%; Pred. No. 7.02e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 yaqgyletrf 87
   :|||:|
QY 8 FGQGYVQTFF 17

RESULT
ID R05999 standard; protein; 361 AA.
AC R05999;
DE 04-DEC-1990 (first entry)
DE P2 gene product of Haemophilus strain 1H.
KW P2 gene; strain 1H; influenza; vaccine; ds.
OS Haemophilus influenzae.
PN EP-378929-A.
PD 25-JUL-1990.
PF 22-DEC-1989; 313573.
PR 23-DEC-1988; GB-030124.
PR 01-FEB-1989; GB-002178.
PA (CONN-) CONNAUGHT LAB LTD.
PI Munson RS, Tolan RW, Chong P, Fahim R, Mcverry P, Klein M;
DR WPI: 90-225607/30.
DR N-PSDB: Q05372.
PT Gene coding for protein P2 of Haemophilus influenzae type-B -
PT used for developing vaccines for protection against disease
PT caused by the organism.
PS Claim 1; Page 8; 15pp; English.
CC Peptide may be used as a vaccine to the disease caused by H1
CC type b, as carrier for conjugation to oligosaccharide derived
CC from Haemophilus. Protein may be used with other haptens as
CC T-cell dependant antigen and carrier.
SQ Sequence 361 AA;

Query Match          64.6%; Score 53; DB 1; Length 361;
Best Local Similarity 50.0%; Pred. No. 7.02e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 yaqgyletrf 87
   :|||:|
QY 8 FGQGYVQTFF 17

RESULT
ID p90098 standard; protein; 362 AA.
AC p90098;
DE 1-NOV-1989 (first entry)
DE P2 antigen of Haemophilus influenzae
KW Haemophilus influenzae P2 antigen; vaccines; pathogenic
KW type b strains.
OS Haemophilus influenzae type b
PN EP-320289-A.
PD 14-JUN-1989.
PF 09-DEC-1988; 311691.
PR 10-DEC-1987; US-249482.
PI (TEXA) Univ of Texas Syst.
PI Hansen EJ;
DR WPI: 89-174562/24.

```

PT DNA encoding Haemophilus influenzae P2 antigen
 PT - used for producing immunogenic compsns. as vaccines
 PT and in diagnostic hybridisation assays.

PS Disclosure; fig 6; 20pp; English.

CC P2 antigen of Haemophilus influenzae (see corresp. N90033).

CC Isolated from type b strains, the P2 antigens are used as vaccines

CC and to detect P2-specific RNA or DNA. Segments used as antigens

CC include amino acids 21-361, 175-197, 260-275, 296-311, 33-353,

CC or their functional equivalents.

SQ Sequence 362 AA;

Query Match 64.68; Score 53; DB 1; Length 362;

Best Local Similarity 50.08; Pred. No. 7.02e+01;

Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 79 yaqgyletrf 88

QY 8 FGQGYVOTPF 17

RESULT 10

ID R66878 standard; Protein; 363 AA.

AC R66878;

DT 22-AUG-1995 (first entry)

DE PNV-2 fusion construct of H. influenzae P2 with T7 gene 10 protein.

CC Haemophilus influenzae type b; outer membrane protein; P2; pNV-2;

KW fusion protein; fusion P-2; gene 10; major capsid protein; T7.

OS Synthetic.

PN W09503069-A.

PD 02-FEB-1995.

PF 22-JUL-1994; 008326.

PR 23-JUL-1993; US-096181.

PA (NANA-) NORTH AMERICAN VACCINE INC.

PI Liang S., Pullen JK, Soper TS, Tai JY;

DR WPI: 95-075024/10.

DR N-PSDB; Q84355.

PT Production of Haemophilus influenzae HibP2 protein - for the

PT prevention of bacterial meningitis in animals

PS Claim 14; Figure 5; 63pp; English.

CC Fusion P2 contains mature outer membrane protein P2 from

CC Haemophilus influenzae type b (hib-P2) with 22 AAs of gene 10 of

CC phase T7 at the N-terminus that were derived from the pET-17b

CC expression vector (Novagen pET System Manual). To P2 into

CC pET-17b the original P2 gene (in pNV-1) (see Q84354) was modified

CC using PCR. To construct the fusion-P2, an oligo was constructed

CC allowed the mature porin to be cloned into the BamHI site of

CC pET-17b, thus yielding a fusion protein P2 to gene 10. The

CC sequence of this oligo was Q84351. The extraneous 3' sequences were

CC eliminated by introducing xhoI site about 40 bp from the translation

CC stop codon. This oligo (Q84352) contd. an xhoI site to allow it to

CC be cloned into the xhoI site of pET-17b. PCR was used to generate a

CC 1.1 kb fragment from the full length P2 (pNV-1) with the 5' oligo

CC contg. the BamHI site and the 3' oligo contg. the xhoI site. This

CC fragment was digested with BamHI and xhoI, purified and ligated

CC into the BamHI-xhoI digested pET-17b. This yielded the fusion-P2

CC construct pNV-2.

SQ Sequence 363 AA;

Query Match 64.68; Score 53; DB 13; Length 363;

Best Local Similarity 50.08; Pred. No. 7.02e+01;

Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 80 yaqgyletrf 89

QY 8 FGQGYVOTPF 17

RESULT 11

ID W21678 standard; Protein; 390 AA.

AC W21678;

DT 30-SEP-1997 (first entry)

DE Haemophilus influenzae outer membrane protein P2.

KW Fusion protein; protein delivery; outer membrane protein P2;

KW vaccine; immunogen; respiratory disease.

OS Haemophilus influenzae strain 1479.

FH Location/Qualifiers

FT region 22..43

FT /label= Loop-1

FT region 69..85

FT /label= Loop-2

FT region 115..133

FT /label= Loop-3

FT region 155..175

FT /label= Loop-4

FT region 203..248

FT /label= Loop-5

FT region 262..277

FT /label= Loop-7

FT region 303..315

FT /label= Loop-7

FT region 331..364

FT /label= Loop-8

PN W09716207-A1.

PD 09-MAY-1997.

PF 01-NOV-1996; U17698.

PR 31-OCT-1996; US-740644.

PR 02-NOV-1995; US-006168.

PA (UYNV) UNIV NEW YORK STATE RES FOUND.

PI Kyungcheol Y, Timothy FM;

DR WPI: 97-271883/24.

DR N-PSDB; T74476.

PT New recombinant DNA for producing fusion proteins - comprising

PT microbial pathogen epitope inserted into Haemophilus influenzae P2

PT protein loop 5, used in vaccine

PS Example 1; Page 40-43; 54pp; English.

CC The loop 5 portion of Haemophilus influenzae strain 1479 outer

CC membrane protein P2 (W21678) provides a structure suitable for

CC fusion with, and display of, one or more heterologous or homologous

CC peptide sequences. A method is provided for producing a fusion

CC protein comprising a P2 protein in which loop 5, or a portion of

CC it, displays a heterologous or homologous peptide sequence which

CC antigenically mimics an epitope from a microbial pathogen. The

CC fusion protein can be expressed in a host cell (e.g. as a surface-

CC exposed sequence) and used as an immunogen against disease caused

CC by H. influenzae and/or the microbial pathogen. It is esp. used

CC to treat infection by respiratory pathogens.

SQ Sequence 390 AA;

Query Match 64.68; Score 53; DB 23; Length 390;

Best Local Similarity 50.08; Pred. No. 7.02e+01;

Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 yaqgyletrf 87

QY 8 FGQGYVOTPF 17

RESULT 12

ID R97616 standard; Protein; 796 AA.

AC R97616;

DT 23-AUG-1996 (first entry)

DE Cephalosporin amidohydrolase.

KW Cephalosporin amidohydrolase;

OS 7-aminocephalosporanic acid; gac gene.

PN W09616174-A2.

PD 30-MAY-1996.

PF 15-NOV-1995; E04487.

PR 17-NOV-1994; GB-023212.

PA (GLAX) GLAXO GROUP LTD.

PI Burr KW, Harrison LA, Illing GT, Maishman NJ, Ramsden M;

DR WPI: 96-268610/27.

DR N-PSDB; T29215.

PT One-step conversion of cephalosporin C into 7-aminocephalosporanic

PT acid - allows environmentally friendly production of

7-amino:cephalosporanic acid for prod. of cephalosporin-like antibiotics
PT Disclosure: Page 32-36; 40pp; English.
PS A cephalosporin amidohydrolase (R97616) was identified as the
CC product of the *gac* gene (T29215) of *Pseudomonas vesicularis* strain
CC B965. The enzyme is capable of converting cephalosporin C and
CC its derivatives to 7-aminocephalosporanic acid and its
CC derivatives, building blocks of semisynthetic cephalosporin
CC antibiotics. The method is cheap, involves simple technology and
CC is environmentally safe. The enzyme is obtd. by incorporation of
CC the *gac* gene into a vector and expression in e.g. *Escherichia coli*.
CC The purified enzyme is pref. immobilised.
SQ Sequence 796 AA;

Query Match 62.2%; Score 51; DB 17; Length 796;
Best Local Similarity 100.0%; Pred. No. 1.13e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 83 fgqgyv 88
| | | | |
QY 8 FGQGYV 13

RESULT 13
ID W52846 standard; Protein: 5069 AA.
AC W52846;
DT 24-JUL-1998 (first entry)
DE A. mediterranei rifamycin synthesis gene cluster fragment protein B.
KW Anycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
KW polyketide synthase; actinomycete; ansamycin.
OS Anycolatopsis mediterranei.
PN WO9807868-A1.
PD 26-FEB-1998.
PR 18-AUG-1997; E04495.
PR 20-AUG-1996; EP-810551.
PA (NOVS) NOVARTIS AG.
PI Engel N, Schupp T, Toupet C;
DR N-PSDB: 98-169172/15.
PT Anycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT to produce rifamycin and rifamycin analogues
PS Claim 6: Page 126-151; 205pp; English.
CC The present sequence represents a Anycolatopsis mediterranei rifamycin
CC synthesis gene cluster ORF B protein from the present invention. The
CC DNA fragment comprises a DNA region involved directly or indirectly
CC in the gene cluster responsible for rifamycin synthesis, including
CC the adjacent DNA regions to the right and left which, by reason of
CC their function in connection with rifamycin biosynthesis, qualify
CC as constituents of this rifamycin gene cluster, and functional
CC fragments, derivatives or constituents of these. The Anycolatopsis
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC for producing rifamycin, rifamycin analogues or precursors. It can also
CC be used for inactivating or modifying genes involved in ansamycin or
CC rifamycin biosynthesis. The DNA can be used for constructing mutant
CC actinomycetes strains from which the natural rifamycin or ansamycin
CC biosynthesis gene cluster has been partly or completely deleted. The
CC DNA fragment can be used for assembling a library of polyketide
CC synthases, which can be used for assembling a library of polyketides.
CC A hybridisation probe of the invention can be used for identifying DNA
CC fragments involved in the biosynthesis of ansamycins.
SQ Sequence 5069 AA;

Query Match 62.2%; Score 51; DB 30; Length 5069;
Best Local Similarity 100.0%; Pred. No. 1.13e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3455 fgqgyv 3460
| | | | |
QY 8 FGQGYV 13

RESULT 14
ID W55964 standard; Protein: 613 AA.

7-amino:cephalosporanic acid for prod. of cephalosporin-like antibiotics
PT Disclosure: Page 32-36; 40pp; English.
PS A cephalosporin amidohydrolase (R97616) was identified as the
CC product of the *gac* gene (T29215) of *Pseudomonas vesicularis* strain
CC B965. The enzyme is capable of converting cephalosporin C and
CC its derivatives to 7-aminocephalosporanic acid and its
CC derivatives, building blocks of semisynthetic cephalosporin
CC antibiotics. The method is cheap, involves simple technology and
CC is environmentally safe. The enzyme is obtd. by incorporation of
CC the *gac* gene into a vector and expression in e.g. *Escherichia coli*.
CC The purified enzyme is pref. immobilised.
SQ Sequence 796 AA;

Query Match 61.0%; Score 50; DB 32; Length 613;
Best Local Similarity 60.0%; Pred. No. 1.43e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

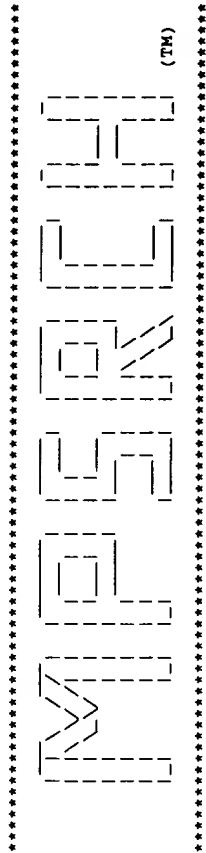
Db 36 fdegghgtgf 45
| | | | |
QY 8 FGQGYVTPF 17

RESULT 15
ID R36774 standard; Protein: 613 AA.
AC R36774;
DT 28-SEP-1993 (first entry)
DE Serratia esterase.
KW S441; M1; mutation; allele; variation; hydrolysis.
OS Serratia marcescens.
PN EP-544250-A.
PD 02-JUN-1993.
PF 25-NOV-1992; 120078.
PR 25-NOV-1991; JP-355440.
PR 13-APR-1992; JP-137502.
PA (TANA) TANABE SEIYAKU CO.
PI Akatsuka H, Komatsubara S, Omori K, Shibatani T;
DR WPI: 93-177043/22.
DR N-PSDB: Q42857, Q42858.
PT New genes encoding esterase from *Serratia* - are contained in
PT recombinant vectors and transformed cells with high esterase
PT productivity, useful as hydrolytic reagent
PS Disclosure: Page 11-13; 16pp; English.
CC When S. marcescens S441 was subjected to mutagenesis with
CC N-methyl-N'-nitro-N-nitrosoguanidine and mutants selected on
CC tributyrin medium, the mutant M1 was identified. M1 contains
CC an allelic variant of *c* -> *t* at position 9. M1 has 2.5 times
CC greater esterase prod. than S441. The esterase is useful for
CC carrying out hydrolytic reactions.
SQ Sequence 613 AA;

Query Match 61.0%; Score 50; DB 7; Length 613;
Best Local Similarity 60.0%; Pred. No. 1.43e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 36 fdegghgtgf 45
| | | | |
QY 8 FGQGYVTPF 17

Search completed: Wed Sep 1 16:23:25 1999
Job time : 25 secs.



Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Sep 1 16:22:24 1999; MasPar time 5.25 Seconds
Tabular output not generated. 76.309 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (8-17) from PCTUS9913024.pep (8 of 12)
Perfect Score: 82
Sequence: 1 FGQGYVQTPE 10

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 24.385; Variance 32.736; scale 0.745

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	57	69.5	262	S64050	hypothetical protein	1.13e+00
2	55	67.1	509	C69491	probable acid-CoA li	2.77e+00
3	55	67.1	1036	A34755	nitrogen regulatory p	2.77e+00
4	54	65.9	547	A43832	heme-binding protein	4.30e+00
5	54	65.9	549	D64098	major outer membrane	6.66e+00
6	53	64.6	359	D64050	major outer membrane	6.66e+00
7	53	64.6	361	A30542	major outer membrane	6.66e+00
8	53	64.6	371	S68072	major outer membrane	6.66e+00
9	53	64.6	371	S68069	major outer membrane	6.66e+00
10	53	64.6	392	C69872	probable aspartate tr	6.66e+00
11	53	64.6	936	B64567	cytochrome c biogenes	6.66e+00
12	53	64.6	936	H71862	probable cytochrome C	6.66e+00
13	52	63.4	60	JN0741	hypothetical 6.6K pro	1.02e+01
14	52	63.4	484	A24994	cellulose 1,4-beta-ce	1.02e+01
15	52	63.4	1123	A44766	defective chorion-1 f	1.02e+01
16	51	62.2	340	G70432	conserved hypothetical	1.57e+01
17	51	62.2	743	S19437	hypothetical protein	1.57e+01
18	51	62.2	822	S26872	DNA-directed RNA poly	1.57e+01
19	51	62.2	830	RNEGB2	hypothetical protein	1.57e+01
20	51	62.2	2301	T02323	hypothetical protein	1.57e+01
21	50	61.0	234	F64595	membrane fusion prote	2.38e+01
22	50	61.0	234	G71918	probable efflux trans	2.38e+01
23	50	61.0	489	C64984	lysine-specific perme	2.38e+01

24	50	61.0	612	2	A54282	reversed polarity pro	2.38e+01
25	50	61.0	613	2	JS0763	triacylglycerol lipas	2.38e+01
26	50	61.0	915	2	T00728	hypothetical protein	3.60e+01
27	49	59.8	336	2	PC1139	hypothetical protein	3.60e+01
28	49	59.8	418	2	F70669	hypothetical protein	3.60e+01
29	49	59.8	425	2	E71039	hypothetical protein	3.60e+01
30	49	59.8	608	2	B64610	outer membrane protei	3.60e+01
31	49	59.8	629	2	E64610	outer membrane protei	3.60e+01
32	49	59.8	638	2	B71904	probable outer membra	3.60e+01
33	49	59.8	651	2	E71904	probable outer membra	3.60e+01
34	49	59.8	699	2	A43734	probable protein kin	3.60e+01
35	49	59.8	779	2	A57177	NIMA-like protein kin	3.60e+01
36	49	59.8	786	2	JC1298	aculeacin-A acylase p	3.60e+01
37	49	59.8	930	2	E64098	DNA-directed DNA poly	3.60e+01
38	48	58.5	241	2	S77548	lumQ protein - Synec	5.41e+01
39	48	58.5	402	2	F64389	chromate resistance p	5.41e+01
40	48	58.5	408	2	C70870	probable lpps protein	5.41e+01
41	48	58.5	481	2	A26483	bindin precursor - se	5.41e+01
42	48	58.5	532	2	S70844	hemolytic protein hhd	5.41e+01
43	48	58.5	625	2	H70330	hypothetical protein	5.41e+01
44	48	58.5	659	2	JC4910	heparitin-sulfate lya	5.41e+01
45	48	58.5	3473	1	A46112	genome polyprotein -	5.41e+01

ALIGNMENTS

RESULT 1
ENTRY S64050 #type complete
TITLE hypothetical protein YGL046w - yeast (Saccharomycetes cerevisiae)
ALTERNATE_NAMES hypothetical protein G3535
ORGANISM #formal_name Saccharomycetes cerevisiae
DATE 17-May-1996 #sequence_revision 17-May-1996 #text_change 14-Nov-1997
ACCESSIONS S64050
REFERENCE S64044
#authors Feuermann, M.; Potier, S.; Souciet, J.L.
#submission submitted to the Protein Sequence Database, May 1996
#accession S64050
#molecule_type DNA
#residues 1-262 ##label FEU
##cross-references EMBL:Z72568; NID:gl322534; PID:e243943; PID:gl322535; MIPS:YGL046w
#experimental_source strain S288C

GENETICS
#map_position 7L
SUMMARY #length 262 #molecular-weight 29440 #checksum 290

Query Match 69.5%; Score 57; DB 2; Length 262;
Best Local Similarity 60.0%; Pred. No. 1.13e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 107 YGQYVKTAF 116
QY 8 FGQGYVQTPE 17
: || || | : |

RESULT 2
ENTRY C69491 #type complete
TITLE probable acid-CoA ligase (EC 6.2.1.-) fadD8 - Archaeoglobus fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 12-Feb-1999
ACCESSIONS C69491
REFERENCE A69250
#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kierlavage, A.R.; Graham, D.E.; Kyripides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.;

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

#journal Nature (1997) 390:364-370
 #title The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*.
 #cross-references MUID:98049343
 #accession C69491
 #status preliminary: nucleic acid sequence not shown; translation not shown

##molecule_type DNA
 ##residues 1-509 ##label KLE
 ##cross-references GB:AE000970; GB:AE000782; NID:g2689293; PID:g2648612; TIGR:AF1932

CLASSIFICATION #superfamily acetate-CoA ligase homology
 KEYWORDS acid-thiol ligase

FEATURE 52-499
 SUMMARY #length 509 #molecular-weight 57035 #checksum 45

Query Match 67.1%; Score 55; DB 2; Length 509;
 Best Local Similarity 60.0%; Pred. No. 2.77e+00;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 325 FTGGYGVVGF 334
 I : I I I I I I I
 QY 8 FGQGYVQTFP 17

RESULT 3
 ENTRY nitrogen regulatory protein nit-2 - *Neurospora crassa*
 TITLE #formal name *Neurospora crassa*
 ORGANISM 13-Jul-1990 #sequence_revision 26-Jul-1996 #text_change 16-Feb-1997

ACCESSIONS A34755
 REFERENCE A34755
 #authors Fu, Y.H.; Marzluf, G.A.
 #journal Mol. Cell. Biol. (1990) 10:1056-1065
 #title nit-2, the major nitrogen regulatory gene of *Neurospora crassa*, encodes a protein with a putative zinc finger DNA-binding domain.
 #cross-references MUID:90158568
 #accession A34755
 ##molecule_type DNA; mRNA
 ##residues 1-1036 ##label FUY
 ##cross-references GB:M33956

GENETICS 209/2; 335/3
 #introns
 CLASSIFICATION #superfamily nitrogen regulatory protein nit-2; GATA-type zinc finger homology
 KEYWORDS DNA binding; transcription regulation; zinc finger
 FEATURE 740-793
 743-767 #domain GATA-type zinc finger homology #label GZF\
 #region zinc finger GATA motif
 SUMMARY #length 1036 #molecular-weight 109295 #checksum 235

Query Match 67.1%; Score 55; DB 1; Length 1036;
 Best Local Similarity 66.7%; Pred. No. 2.77e+00;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 981 FGNGYINTP 989
 I I I I I I I
 QY 8 FGQGYVQTP 16

RESULT 4
 ENTRY heme-binding protein A precursor - *Haemophilus influenzae*
 TITLE (type b)

ORGANISM #formal name *Haemophilus influenzae*
 DATE 07-May-1993 #sequence_revision 07-May-1993 #text_change 17-Oct-1997

ACCESSIONS A43832
 REFERENCE A43832
 #authors Hanson, M.S.; Slaughter, C.; Hansen, E.J.
 #journal Infect. Immun. (1992) 60:2257-2266
 #title The hbpA gene of *Haemophilus influenzae* type b encodes a heme-binding lipoprotein conserved among heme-dependent *Haemophilus* species.

#cross-references MUID:92267636
 #contents type b
 #accession A43832
 ##molecule_type DNA; protein
 ##residues 1-547 ##label HAN
 ##cross-references GB:M88134
 ##note sequence extracted from NCBI backbone (NCBIN:104242, NCBIP:104243)

GENETICS
 #gene hbpA
 CLASSIFICATION #superfamily dipeptide transport protein
 KEYWORDS heme
 FEATURE 1-18
 19-547 #domain signal sequence #status predicted #label SIG\
 #product heme-binding protein A #status predicted #label MAT

SUMMARY #length 547 #molecular-weight 60692 #checksum 7908
 Query Match 65.9%; Score 54; DB 2; Length 547;
 Best Local Similarity 85.7%; Pred. No. 4.30e+00;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 528 GYQSPF 534
 I I I I I I I
 QY 11 GYVQTFP 17

RESULT 5
 ENTRY heme-binding protein A precursor - *Haemophilus influenzae*
 TITLE (strain Rd K20)

ORGANISM #formal name *Haemophilus influenzae*
 DATE 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 24-Oct-1997

ACCESSIONS D64098
 REFERENCE A64000
 #authors Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.; FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
 #journal Science (1995) 269:496-512
 #title Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.
 #cross-references MUID:95350630
 #accession D64098
 ##status nucleic acid sequence not shown; translation not shown
 ##molecule_type DNA
 ##residues 1-549 ##label TIGR
 ##cross-references GB:U32767; GB:L42023; NID:g1573868; PID:g1573869; TIGR:H10853

GENETICS
 #gene hbpA
 CLASSIFICATION #superfamily dipeptide transport protein
 KEYWORDS heme
 FEATURE 1-20
 21-549 #domain signal sequence #status predicted #label SIG\
 #product heme-binding protein A #status predicted #label


```

SUMMARY      MAT      #length 549  #molecular-weight 60904  #checksum 3028
Query Match   65.9%; Score 54; DB 2; Length 549;
Best Local Similarity 85.7%; Pred. No. 4.30e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 530 GYVQSPF 536
      :::::|
QY 11 GYVQTPF 17

RESULT      6
ENTRY      D64050  #type complete
TITLE      major outer membrane protein P2 - Haemophilus influenzae
           (strain Rd KW20)
ORGANISM   #formal_name Haemophilus influenzae
DATE       18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
           24-Oct-1997
ACCESSIONS D64050
REFERENCE   A64000
#authors   Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
           Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.;
           Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
           Fitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
           Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
           J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
           M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
           D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,
           J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
           Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
#journal   Science (1995) 269:496-512
#title     Whole-genome random sequencing and assembly of Haemophilus
           influenzae Rd.
#cross-references MUID:95350630
#accession D64050
#status    nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues  1-359 ##label TIGR
#cross-references GB:U32699; GB:L42023; NID:g1573084; PID:g1573092;
           TIGR:H10139
KEYWORDS   membrane protein
SUMMARY    #length 359  #molecular-weight 39375  #checksum 1781

Query Match   64.6%; Score 53; DB 2; Length 359;
Best Local Similarity 50.0%; Pred. No. 6.66e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 YAOGYLETRF 87
      :::::|
QY 8 FGQGVQTPF 17

RESULT      7
ENTRY      A30542  #type complete
TITLE      major outer membrane protein P2 precursor - Haemophilus
           influenzae (type b)
ORGANISM   #formal_name Haemophilus influenzae
DATE       07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
           20-Mar-1998
ACCESSIONS A30542; A32750; A60116; A60121; S09622
REFERENCE   A30542
#authors   Munson Jr., R.; Tolan Jr., R.W.
#journal   Infect. Immun. (1989) 57:88-94
#title     Molecular cloning, expression, and primary sequence of outer
           membrane protein P2 of Haemophilus influenzae type b.
#cross-references MUID:89079316
#accession A30542
#molecule_type DNA
#residues  1-361 ##label MUN
#cross-references GB:J03359; NID:g148958; PID:g148959
#experimental_source type b
#note      part of this sequence, including the amino end of the

```

```

REFERENCE   A32750  mature protein, was confirmed by protein sequencing
#authors   Hansen, E.J.; Hasemann, C.; Clause, A.; Capra, J.D.; Orth,
           K.; Moomaw, C.R.; Slaughter, C.A.; Latimer, J.L.; Miller,
           E.E.
#journal   Infect. Immun. (1989) 57:1100-1107
#title     Primary structure of the porin protein of Haemophilus
           influenzae type b determined by nucleotide sequence
           analysis.
#cross-references MUID:89173305
#accession A32750
#status    not compared with conceptual translation
#molecule_type DNA
#residues  1-361 ##label HAN
#experimental_source type b, strain DL42
#note      mature protein, was confirmed by protein sequencing

REFERENCE   A60116
#authors   Hansen, E.J.; Pelzel, S.E.; Orth, K.; Moomaw, C.R.; Radolf,
           J.D.; Slaughter, C.A.
#journal   Infect. Immun. (1989) 57:3270-3275
#title     Structural and antigenic conservation of the P2 porin protein
           among strains of Haemophilus influenzae type b.
#cross-references MUID:90035390
#accession A60116
#molecule_type protein
#residues  21-42 ##label HA2
#experimental_source type b, strains DL42, CH100, DL26, DV102, H234,
           Madigan, OA104, and SL103

REFERENCE   A60121
#authors   Hansen, E.J.; Gonzales, F.R.; Chamberlain, N.R.; Norgard,
           M.V.; Miller, E.E.; Cope, L.D.; Pelzel, S.E.; Gaddy, B.;
           Claussell, A.
#journal   Infect. Immun. (1988) 56:2709-2716
#title     Cloning of the gene encoding the major outer membrane protein
           of Haemophilus influenzae type b.
#cross-references MUID:88330197
#accession A60121
#molecule_type protein
#residues  21-40 ##label HA3
#experimental_source type b, strain DL42

REFERENCE   S09622
#authors   Munson Jr., R.; Bailey, C.; Grass, S.
#journal   Mol. Microbiol. (1989) 3:1797-1803
#title     Diversity of the outer membrane protein P2 gene from major
           clones of Haemophilus influenzae type b.
#cross-references MUID:90158127
#accession S09622
#status    preliminary; not compared with conceptual translation
#molecule_type DNA
#residues  1-361 ##label MU2
KEYWORDS   membrane protein
FEATURE    1-20
           21-361
SUMMARY    #domain signal sequence #status predicted #label SIG\
           #product major outer membrane protein P2 #status
           experimental #label MAT
           #length 361 #molecular-weight 39701 #checksum 4550
Query Match   64.6%; Score 53; DB 2; Length 361;
Best Local Similarity 50.0%; Pred. No. 6.66e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 YAOGYLETRF 87
      :::::|
QY 8 FGQGVQTPF 17

RESULT      8
ENTRY      S68072  #type complete
TITLE      major outer membrane protein P2, antigenic variant t1 -
           Haemophilus influenzae
ORGANISM   #formal_name Haemophilus influenzae
DATE       19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change

```

```

24-Oct-1997
S68072: S45437
REFERENCE
#authors
#submission
#accession
#status
#molecule_type DNA
#residues 1-371 #label DU1
#cross-references EMBL:X73386; NID:g510590; PID:g510591
#experimental_source isolate t1
REFERENCE
S43699
#authors
#journal
#title
Mol. Microbiol. (1994) 11:1181-1189
Antigenic drift of non-encapsulated Haemophilus influenzae
major outer membrane protein p2 in patients with chronic
bronchitis is caused by point mutations.
#cross-references MUID:94293786
#accession S45437
#status preliminary
#molecule_type DNA
#residues 174-188;215-234;257-283 #label DUW
#cross-references EMBL:X73386
#experimental_source isolate t1
KEYWORDS
membrane protein
SUMMARY
#length 371 #molecular-weight 40901 #checksum 3994
Query Match 64.6%; Score 53; DB 2; Length 371;
Best Local Similarity 50.0%; Pred. No. 6.66e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 YAOGYLETRF 87
:||||:|
QY 8 FGQGYVQTF 17

RESULT 9
ENTRY
TITLE
ORGANISM
DATE
#status
#molecule_type DNA
#residues 1-371 #label DU1
#cross-references EMBL:X73383; NID:g475192; PID:e1192128; PID:g2654292
#experimental_source isolate t2
REFERENCE
S43699
#authors
#journal
#title
Mol. Microbiol. (1994) 11:1181-1189
Antigenic drift of non-encapsulated Haemophilus influenzae
major outer membrane protein p2 in patients with chronic
bronchitis is caused by point mutations.
#cross-references MUID:94293786
#accession S43706
#status preliminary
#molecule_type DNA
#residues 174-188;215-234;257-283 #label DUW
#cross-references EMBL:X73383
#experimental_source isolate t2
KEYWORDS
membrane protein
SUMMARY
#length 371 #molecular-weight 40945 #checksum 5124
Query Match 64.6%; Score 53; DB 2; Length 371;
Best Local Similarity 50.0%; Pred. No. 6.66e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

24-Oct-1997
S68072: S45437
REFERENCE
#authors
#submission
#accession
#status
#molecule_type DNA
#residues 1-371 #label DU1
#cross-references EMBL:X73386; NID:g510590; PID:g510591
#experimental_source isolate t1
REFERENCE
S43699
#authors
#journal
#title
Mol. Microbiol. (1994) 11:1181-1189
Antigenic drift of non-encapsulated Haemophilus influenzae
major outer membrane protein p2 in patients with chronic
bronchitis is caused by point mutations.
#cross-references MUID:94293786
#accession S45437
#status preliminary
#molecule_type DNA
#residues 174-188;215-234;257-283 #label DUW
#cross-references EMBL:X73386
#experimental_source isolate t1
KEYWORDS
membrane protein
SUMMARY
#length 371 #molecular-weight 40901 #checksum 3994
Query Match 64.6%; Score 53; DB 2; Length 371;
Best Local Similarity 50.0%; Pred. No. 6.66e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 YAOGYLETRF 87
:||||:|
QY 8 FGQGYVQTF 17

RESULT 10
ENTRY
TITLE
ORGANISM
DATE
#status
#molecule_type DNA
#residues 1-392 #label KUN
#cross-references GB:AL009126; NID:g2633699; PID:e1184990;
#experimental_source strain 168
REFERENCE
A33496
#authors
#journal
#title
J. Bacteriol. (1989) 171:6187-6196
Characterization of the gene for a protein kinase which
phosphorylates the sporulation-regulatory proteins spo0A
and spo0F of Bacillus subtilis.
#cross-references MUID:90036708
#accession C33496
#status
#molecule_type DNA
#residues 1-96,'S',99-232,'T',234-392 #label PER
#experimental_source strain 168

```

```

Db 78 YAOGYLETRF 87
:||||:|
QY 8 FGQGYVQTF 17

RESULT 10
ENTRY
TITLE
ORGANISM
DATE
#status
#molecule_type DNA
#residues 1-392 #label KUN
#cross-references GB:AL009126; NID:g2633699; PID:e1184990;
#experimental_source strain 168
REFERENCE
A33496
#authors
#journal
#title
J. Bacteriol. (1989) 171:6187-6196
Characterization of the gene for a protein kinase which
phosphorylates the sporulation-regulatory proteins spo0A
and spo0F of Bacillus subtilis.
#cross-references MUID:90036708
#accession C33496
#status
#molecule_type DNA
#residues 1-96,'S',99-232,'T',234-392 #label PER
#experimental_source strain 168

C69672 #type complete
probable aspartate transaminase (EC 2.6.1.1) - Bacillus
subtilis
#formal_name Bacillus subtilis
#formal_name Bacillus subtilis
05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
16-Oct-1998
A69580
Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Biolotin, A.; Borchert, S.; Boriss, R.G.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Erington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moesti, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porvolik, S.; Prescott,
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Seror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, B.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Banchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession C69672
#status
#molecule_type DNA
#residues 1-392 #label KUN
#cross-references GB:AL009126; NID:g2633699; PID:e1184990;
#experimental_source strain 168
REFERENCE
A33496
#authors
#journal
#title
J. Bacteriol. (1989) 171:6187-6196
Characterization of the gene for a protein kinase which
phosphorylates the sporulation-regulatory proteins spo0A
and spo0F of Bacillus subtilis.
#cross-references MUID:90036708
#accession C33496
#status
#molecule_type DNA
#residues 1-96,'S',99-232,'T',234-392 #label PER
#experimental_source strain 168

```

```

GENETICS
#gene      pata
CLASSIFICATION
#superfamily aspartate transaminase
#amino transferase
KEYWORDS
SUMMARY    #length 392 #molecular-weight 43448 #checksum 5874

Query Match      64.6%; Score 53; DB 2; Length 392;
Best Local Similarity 50.0%; Pred. No. 6.66e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 351 YGEGYVRLSF 360
      :|::|::|:
QY 8 FGQGYVQTFF 17

RESULT 11
ENTRY   B64567 #type complete
TITLE   cytochrome c biogenesis protein - Helicobacter pylori (strain
26695)
ORGANISM #formal_name Helicobacter pylori
DATE     09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
10-Oct-1997
ACCESSIONS B64567
REFERENCE   A64520
#authors    Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
Richardson, D.; Dodson, R.; Khalaf, H.G.; Glodek, A.;
McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
C.M.; Venter, J.C.

#journal    Nature (1997) 388:539-547
#title      The complete genome sequence of the gastric pathogen
Helicobacter pylori.
#cross-references GB:AE00554; GB:AE00554; NID:92313475; PID:92313481;
#accession  B64567
#status     preliminary: nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues   1-936 #label TOM
#cross-references GB:AE00554; GB:AE00554; NID:92313475; PID:92313481;
TIGR:HP0378
SUMMARY     #length 936 #molecular-weight 106359 #checksum 8931

Query Match      64.6%; Score 53; DB 2; Length 936;
Best Local Similarity 50.0%; Pred. No. 6.66e+00;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 222 FGSAYIELPF 231
      ||::|::|
QY 8 FGQGYVQTFF 17

RESULT 12
ENTRY   H71862 #type complete
TITLE   probable cytochrome C-type biogenesis protein - Helicobacter
pylori (strain J99)
ORGANISM #formal_name Helicobacter pylori
#variety  strain J99
DATE     12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
12-Feb-1999
ACCESSIONS H71862
REFERENCE   A71800
#authors    Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
G.F.; Trust, T.J.

```

```

#journal    Nature (1999) 397:176-180
#title      Genomic sequence comparison of two unrelated isolates of the
human gastric pathogen Helicobacter pylori.
#cross-references MUID:99120557
#accession  H71862
#status     preliminary
#molecule_type DNA
#residues   1-936 #label ARN
#cross-references GB:AE001528; GB:AE001439; NID:94155575; PID:94155589
#experimental_source strain J99

GENETICS
#gene      jhp1003
SUMMARY     #length 936 #molecular-weight 106452 #checksum 7770

Query Match      64.6%; Score 53; DB 2; Length 936;
Best Local Similarity 50.0%; Pred. No. 6.66e+00;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 222 FGSAYIELPF 231
      ||::|::|
QY 8 FGQGYVQTFF 17

RESULT 13
ENTRY   JN0741 #type complete
TITLE   hypothetical 6.6K protein - phage SPp1
ALTERNATE_NAMES hypothetical protein 50
ORGANISM #formal_name phage SPp1
DATE     03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change
17-Mar-1999
ACCESSIONS JN0741; S21439
REFERENCE   JN0729
#authors    Chai, S.; Szepan, U.; Lueder, G.; Trautner, T.A.; Alonso,
J.C.

#journal    Gene (1993) 129:41-49
#title      Sequence analysis of the left end of the Bacillus subtilis
bacteriophage SPp1 genome.
#cross-references EMBL:X65941; NID:914843; PID:g579088
#accession  JN0741
#molecule_type DNA
#residues   1-60 #label CHA
#cross-references EMBL:X65941; NID:914843; PID:g579088
SUMMARY     #length 60 #molecular-weight 6692 #checksum 7702

Query Match      63.4%; Score 52; DB 2; Length 60;
Best Local Similarity 50.0%; Pred. No. 1.02e+01;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 28 YNGYIKTGF 37
      :|::|:
QY 8 FGQGYVQTFF 17

RESULT 14
ENTRY   A24994 #type complete
TITLE   cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) A precursor -
Cellulomonas fimi
ALTERNATE_NAMES exo-cellobiohydrolase A
ORGANISM #formal_name Cellulomonas fimi
DATE     15-Dec-1988 #sequence_revision 22-Nov-1996 #text_change
10-Jul-1998
ACCESSIONS A24994
REFERENCE   A24994
#authors    O'Neill, G.; Goh, S.H.; Warren, R.A.J.; Kilburn, D.G.; Miller
Jr., R.C.
#journal    Gene (1986) 44:325-330
#title      Structure of the gene encoding the exoglucanase of
Cellulomonas fimi.
#cross-references MUID:87055250
#accession  A24994
#molecule_type DNA
#residues   1-484 #label ONE
#cross-references GB:M15824; NID:g144424; PID:g144425

```

##note the amino-terminal sequence of the mature protein (residues 42-71) has been determined (M.L. Langsford, unpublished data)

REFERENCE A67217 White, A.; Withers, S.G.; Gilkes, N.R.; Rose, D.R. submitted to the Brookhaven Protein Data Bank, July 1994

##cross-references PDB:2EXO

##contents annotation: X-ray crystallography, 1.8 angstroms, residues 42-173, 'G', 175, 'RR', 178-353

REFERENCE A55905 White, A.; Withers, S.G.; Gilkes, N.R.; Rose, D.R. Biochemistry (1994) 33:12546-12552

##authors

##journal Crystal structure of the catalytic domain of the beta-1,4-glycanase Cex from Cellulomonas fimi.

##contents annotation: X-ray crystallography, 1.8 angstroms, residues 42-173, 'G', 175, 'RR', 178-353

GENETICS

##gene cex

##description hydrolyses 1,4-beta-D-glucosidic linkages in cellulose and, releasing cellobiose from the non-reducing ends of the chains

CLASSIFICATION #superfamily Cellulomonas cellulose 1,4-beta-cellobiosidase A; bacterial cellulose-binding domain homology; Streptomyces endo-1,4-beta-xylanase A homology; extracellular protein; glycoprotein; glycosidase; hydrolase; polysaccharide degradation; tandem repeat

KEYWORDS

FEATURE

1-41 #domain signal sequence #status predicted #label SIG\

42-484 #product cellulose 1,4-beta-cellobiosidase #status experimental #label MAT\

73-353 #domain streptomyces endo-1,4-beta-xylanase A homology #label SX\

357-377 #region 7-residue repeats (P-T-P-T-P-T-[S/T])\

381-482 #domain bacterial cellulose-binding domain homology #label BCB\

137,398,403,447, #binding_site carbohydrate (Asn) (covalent) #status predicted\

461

168,274 #active_site Glu #status experimental\

208-240,302-308 #disulfide_bonds #status experimental\

382-481 #disulfide_bonds #status predicted

SUMMARY #length 484 #molecular-weight 51291 #checksum 9660

Query Match 63.4%; Score 52; DB 1; Length 484;
Best Local Similarity 50.0%; Pred. No. 1.02e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 186 LGNGYIETAF 195
:| |||:|
Qy 8 FGQGYVQTPF 17

RESULT 15

ENTRY A44766 #type complete

TITLE defective chorion-1 fci25 protein precursor - fruit fly (Drosophila melanogaster)

ORGANISM #formal_name Drosophila melanogaster

DATE 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 24-Sep-1998

ACCESSIONS A44766

REFERENCE A44766

##authors Waring, G.L.; Hawley, R.J.; Schoenfeld, T.

##journal Dev. Biol. (1990) 142:1-12

##title Multiple proteins are produced from the dec-1 eggshell gene in Drosophila by alternative RNA splicing and proteolytic cleavage events.

##cross-references MUID:91032553

##accession A44766

##status preliminary

##molecule_type mRNA

##residues 1-1123 #label WAR

##cross-references GB:M35887; NID:gl57181; PID:gl57182

GENETICS

##gene FlyBase:dec-1

##cross-references FlyBase:FBgn0000427

SUMMARY #length 1123 #molecular-weight 127959 #checksum 6533

Query Match 63.4%; Score 52; DB 2; Length 1123;
Best Local Similarity 50.0%; Pred. No. 1.02e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1098 FIOGYAESPY 1107
| ||| :|
Qy 8 FGQGYVQTPF 17

Search completed: Wed Sep 1 16:22:40 1999
Job time : 16 secs.

Result No.	Score	Query		Length	DB	ID	Description	Pred. No.
		Match	%					
1	57	69.5	262	1	YGE6_YEAST	1	HYPOTHETICAL 29.4 KD P	3.49e-01
2	55	67.1	1036	1	NT12_NEUCR	1	NITROGEN CATABOLIC ENZ	9.37e+01
3	54	65.9	547	1	HBP4_HAEIN	1	HEME-BINDING PROTEIN A	1.52e+00
4	53	64.6	359	1	OM21_HAEIN	1	OUTER MEMBRANE PROTEIN	2.46e+00
5	53	64.6	361	1	OM22_HAEIN	1	OUTER MEMBRANE PROTEIN	2.46e+00
6	53	64.6	371	1	OM25_HAEIN	1	OUTER MEMBRANE PROTEIN	2.46e+00
7	53	64.6	392	1	PATA_BACSU	1	PUTATIVE AMINOTRANSFER	2.46e+00
8	52	63.4	484	1	GUX_CELFI	1	EXOGLUCANASE PRECURSOR	3.94e+00
9	52	63.4	1123	1	DC11_DROME	1	DEFECTIVE CHORION-1 FC	3.94e+00
10	51	62.2	415	1	R13_DROME	1	60S RIBOSOMAL PROTEIN	6.29e+00
11	51	62.2	743	1	YCR6_YEAST	1	HYPOTHETICAL 84.9 KD P	6.29e+00
12	51	62.2	830	1	RPOD_EUGR	1	DNA-DIRECTED RNA POLYM	6.29e+00
13	50	61.0	488	1	LYSP_ECOLI	1	LYSINE-SPECIFIC PERMEAS	9.95e+00
14	49	59.8	115	1	TR10_HUMAN	1	THYROID RECEPTOR INTER	1.56e+01
15	49	59.8	381	1	DHB2_MOUSE	1	ESTRADIOL 17 BETA-DEHY	1.56e+01
16	49	59.8	699	1	NIMA_EMENI	1	G2-SPECIFIC PROTEIN KI	1.56e+01
17	49	59.8	722	1	GUNF_CLOCE	1	ENDOGLUCANASE F PRECUR	1.56e+01
18	49	59.8	779	1	NIM1_NEUCR	1	G2-SPECIFIC PROTEIN KI	1.56e+01
19	49	59.8	786	1	AAC_ACTUT	1	ACULEXIN A ACYLASE PR	1.56e+01
20	49	59.8	930	1	DPOL_HAEIN	1	DNA POLYMERASE I (EC 2	1.56e+01
21	49	59.8	942	1	DPOL_CHLAU	1	DNA POLYMERASE I (EC 2	1.56e+01
22	48	58.5	381	1	DHB2_RAT	1	ESTRADIOL 17 BETA-DEHY	2.44e+01
23	48	58.5	402	1	Y718_METJA	1	HYPOTHETICAL PROTEIN M	2.44e+01

01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DT NITROGEN CATABOLIC ENZYME REGULATORY PROTEIN.
 DE NIT-2
 GN NEUROSPORA CRASSA.
 OS EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
 OC SORDARIALES; SORDARIACEAE; NEUROSPORA.
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=74-OR23-1A;
 RC MEDLINE: 90158568.
 RX FU Y.-H., MARZLUF G.A.;
 RA "nit-2, the major nitrogen regulatory gene of *Neurospora crassa*,
 RT encodes a protein with a putative zinc finger DNA-binding domain.";
 RL MOL. CELL. BIOL. 10:1056-1065(1990).
 [2]
 RN MUTAGENESIS.
 RP MEDLINE: 91186820.
 RX FU Y.-H., MARZLUF G.A.;
 RA "Site-directed mutagenesis of the 'zinc finger' DNA-binding domain of
 RT the nitrogen-regulatory protein Nit2 of *Neurospora*.";
 RL MOL. MICROBIOL. 4:1847-1852(1990).
 CC -!- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. DURING CONDITIONS
 OF NITROGEN LIMITATION IT TURNS ON THE EXPRESSION OF GENES FOR
 CC ENZYMES WHICH ARE REQUIRED FOR THE USE OF A VARIETY OF SECONDARY
 CC NITROGEN SOURCES, INCLUDING NITRATES, PURINES, AMINO ACIDS, AND
 CC PROTEINS.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- INDUCTION: BY LACK OF A PRIMARY NITROGEN SOURCE.
 CC -!- SIMILARITY: HIGH TO OTHER FUNGAL NITROGEN REGULATORY PROTEINS.
 CC -!- SIMILARITY: TO GATA TRANSCRIPTION FACTORS IN THE ZINC-FINGER
 CC REGION.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M33956; G1107471; .
 DR PIR: A34755; A34755.
 DR PROSITE: PS00344; GATA_ZN_FINGER; 1.
 DR PFAM: PF00320; GATA; 1.
 DR HSP: P17429; 5GAT; 1.
 DR TRANSFAC: T00627; .
 KW TRANSCRIPTION REGULATION; ACTIVATOR: DNA-BINDING; ZINC-FINGER;
 FT NUCLEAR PROTEIN; NITRATE ASSIMILATION; REPEAT.
 FT DOMAIN 49 110 3 X APPROXIMATE REPEATS.
 FT REPEAT 49 55 1.
 FT REPEAT 87 92 2.
 FT REPEAT 105 110 3.
 FT ZN_FING 743 767 GATA-TYPE.
 FT MUTAGEN 743 746 CTNC->STNG: ABOLISH DNA-BINDING.
 FT MUTAGEN 755 756 RR->GG: ABOLISH DNA-BINDING.
 FT MUTAGEN 765 766 NA->DV: ABOLISH DNA-BINDING.
 FT MUTAGEN 768 769 GL->DV: ABOLISH DNA-BINDING.
 FT MUTAGEN 789 790 KR->NS: ABOLISH DNA-BINDING.
 SQ SEQUENCE 1036 AA; 109310 MW; CCDE5BE8 CRC32;
 Query Match 67.1%; Score 55; DB 1; Length 1036;
 Best Local Similarity 66.7%; Pred. No. 9.37e-01;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Db 981 FGNQYINTP 989
 QY 8 FGGYVQTP 16
 RESULT 3
 ID HBPA_HAEIN STANDARD; PRT; 547 AA.
 AC P33950;
 DT 01-FEB-1994 (REL. 28, CREATED)

01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT HEME-BINDING PROTEIN A PRECURSOR (HEMIN-BINDING LIPOPROTEIN).
 DE HBPA OR DPPA OR H10853.
 GN HAEMOPHILUS INFLUENZAE.
 OS BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURILLACEAE;
 OC HAEMOPHILUS.
 [1]
 RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RP STRAIN-DL42 / SEROTYPE B;
 RC MEDLINE: 92267636.
 RX HANSON M.S., SLAUGHTER C., HANSEN E.J.;
 RA "The hbpa gene of *Haemophilus influenzae* type b encodes a
 RT heme-binding lipoprotein conserved among heme-dependent *Haemophilus*
 RL species.";
 RL INFECT. IMMUN. 60:2257-2266(1992).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN-RD / KW20;
 RC MEDLINE: 95350630.
 RX FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
 RA KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
 RA SCOTTNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
 RA MCNETT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
 RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
 RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
 RA FINE L.D., FRITZMAN J.L., FUHRMANN J.L., GEORGEHAGEN N.S.M.,
 RA GNEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
 RA VENTER J.C.;
 RT "Whole-genome random sequencing and assembly of *Haemophilus*
 RT *influenzae* Rd.";
 RL SCIENCE 269:496-512(1995).
 [3]
 RN POSSIBLE FUNCTION, AND SUBCELLULAR LOCATION.
 RP STRAIN-DL42 / SEROTYPE B;
 RC MEDLINE: 91251755.
 RX HANSON M.S., HANSEN E.J.;
 RA "Molecular cloning, partial purification, and characterization of a
 RT heme-binding lipoprotein from *Haemophilus influenzae* type b.";
 RL MOL. MICROBIOL. 5:267-278(1991).
 CC -!- FUNCTION: IMPORTANT ROLE IN HEME ACQUISITION OR METABOLISM.
 CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE INNER MEMBRANE BY A LIPID
 CC ANCHOR.
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
 CC PROTEIN FAMILY 5.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M8134; G148941; ALT_SEQ.
 DR EMBL: M84028; G148894; .
 DR EMBL: U32767; G1573869; ALT_INIT.
 DR PIR: A43832; A43832.
 DR TIGR: H10853; .
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 DR PROSITE: PS01040; SBP_BACTERIAL_5; 1.
 DR PFAM: PF00496; SBP_Bac_5; 1.
 DR HSP: P23847; 1DPP.
 KW INNER MEMBRANE; SIGNAL; LIPOPROTEIN.
 FT SIGNAL 1 18 PROBABLE.
 FT CHAIN 19 547
 FT LIPID 19 19 N-ACYL DIGLYCERIDE (PROBABLE).
 FT VARIANT 48 49 KA -> NS (IN STRAIN DL42).
 FT VARIANT 181 181 T -> N (IN STRAIN DL42).
 FT VARIANT 240 240 H -> N (IN STRAIN DL42).
 FT VARIANT 343 343 T -> I (IN STRAIN DL42).
 FT VARIANT 375 375 A -> V (IN STRAIN DL42).
 SQ SEQUENCE 547 AA; 60660 MW; 5220E11D CRC32;

Query Match 65.9%; Score 54; DB 1; Length 547;
Best Local Similarity 85.7%; Pred. No. 1.52e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 528 GYVQSPF 534
|||||
QY 11 GYVQTPF 17

RESULT 4

ID OM21_HAEIN STANDARD; PRT; 359 AA.
AC P43839;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2 OR H10139.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20;
RX MEDLINE; 95350630.
RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., COTTON M.D.,
RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA VENTER J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd.";
RL Influenzae Rd.";
RL SCIENCE 269:496-512(1995).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -!- SIMILARITY: TO PHOTOBACTERIUM STRAIN SS9 OMPH.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U32699; G1573092; -
CC TIGR: H10139; -
DR OUTER MEMBRANE; TRANSMEMBRANE; PORIN; SIGNAL.
KW SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 359 OUTER MEMBRANE PROTEIN P2.
FT CHAIN 21 359 OUTER MEMBRANE PROTEIN P2.
SQ SEQUENCE 359 AA; 39375 MW; 40615FB6 CRC32;

Query Match 64.6%; Score 53; DB 1; Length 359;
Best Local Similarity 50.0%; Pred. No. 2.46e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

DB 78 YAGGYLETRF 87
:|||||
QY 8 FGQGYVQTPF 17

RESULT 5

ID OM22_HAEIN STANDARD; PRT; 361 AA.
AC P20149;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2.
OS HAEMOPHILUS INFLUENZAE.

OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T1, AND T2;
RX MEDLINE; 94293786.
RA DUIM B., VAN ALPHEN L., EIJK P., JANSEN H.M., DANKERT J.;
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major

OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE B;
RX MEDLINE; 89079316.
RA MONSON R.S. JR., TOLAN R.W. JR.;
RT "Molecular cloning, expression, and primary sequence of outer
membrane protein P2 of Haemophilus influenzae type b.";
RL INFECT. IMMUN. 57:88-94(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE B;
RX MEDLINE; 90158127.
RA MONSON R.S. JR., BAILEY C., GRASS S.;
RT "Diversity of the outer membrane protein P2 gene from major clones of
Haemophilus influenzae type b.";
RL MOL. MICROBIOL. 3:1797-1803(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROTYPE B;
RX MEDLINE; 89173305.
RA HANSEN E.J., HASEMANN C., CLAUSELL A., CAPRA J.D., ORTH K.,
RA MOOMAW C.R., SLAUGHTER C.A., LATIMER J.L., MILLER E.E.;
RT "Primary structure of the porin protein of Haemophilus influenzae
type b determined by nucleotide sequence analysis.";
RL INFECT. IMMUN. 57:1100-1107(1989).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -!- SIMILARITY: TO PHOTOBACTERIUM STRAIN SS9 OMPH.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J03359; G148959; -
CC PIR: A09003; G412251; -
DR PIR; A30542; A30542.
DR PIR; S09622; S09622.
KW OUTER MEMBRANE; TRANSMEMBRANE; PORIN; SIGNAL.
FT SIGNAL 1 20
FT CHAIN 21 361 OUTER MEMBRANE PROTEIN P2.
SQ SEQUENCE 361 AA; 39701 MW; 291391ED CRC32;

Query Match 64.6%; Score 53; DB 1; Length 361;
Best Local Similarity 50.0%; Pred. No. 2.46e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

DB 78 YAGGYLETRF 87
:|||||
QY 8 FGQGYVQTPF 17

RESULT 6

ID OM25_HAEIN STANDARD; PRT; 371 AA.
AC P46027;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T1, AND T2;
RX MEDLINE; 94293786.
RA DUIM B., VAN ALPHEN L., EIJK P., JANSEN H.M., DANKERT J.;
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major

RX MEDLINE; 98400502.
RA NOTENBOOM V., BIRSAN C., NITZ M., ROSE D.R., WARREN R.A.,
RA WITHERS S.G.;
RT "Insights into transition state stabilization of the beta-1,4-
RT glycosidase Cex by covalent intermediate accumulation in active site
RT mutants";
RL NAT. STRUCT. BIOL. 5:812-818(1998).
RN [7]
RP STRUCTURE BY NMR OF 377-484.
RX MEDLINE; 95284032.
RA XU G.-Y., ONG E., GILKES N.R., KILBURN D.G., MUHANDIRAM D.R.,
RA HARRIS-BRANDTS M., CARVER J.P., KAY L.E., HARVEY T.S.;
RT "Solution structure of a cellulose-binding domain from Cellulomonas
RT fimi by nuclear magnetic resonance spectroscopy.";
RL BIOCHEMISTRY 34:6993-7009(1995).
RN [8]
RP MUTAGENESIS OF GLU-168.
RX MEDLINE; 94250681.
RA MACLEOD A.M., LINDHORST T., WITHERS S.G., WARREN R.A.J.;
RT "The acid/base catalyst in the exoglucanase/xyylanase from
RT Cellulomonas fimi is glutamic acid 127: evidence from detailed
RT kinetic studies of mutants.";
RL BIOCHEMISTRY 33:6371-6376(1994).
RN [8]
RP FUNCTION: HYDROLYSES BOTH CELLULOSE AND XYLAN. HAS ALSO WEAK
CC ENDOGLUCANASE ACTIVITY.
CC -!- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC (2) EXOCELLULOHYDROLASES THAT CUT THE DISACCHARIDE CELLULOSE
CC (3) FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES
CC IN CELLULOSE AND CELLOTETRAOSE, RELEASING CELLULOSE FROM THE NON-
CC REDUCING ENDS OF THE CHAINS.
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-XYLOSIDIC
CC LINKAGES IN XYLANS.
CC -!- THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL SITE
CC FOR PROTEOLYSIS.
CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY F (FAMILY 10 OF GLYCOSYL
CC HYDROLASES).
CC -!- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M15824; G144425; -
CC PIR; A24994; A24994.
CC PDB; 2EXO; 08-MAR-95.
CC PDB; 1EXG; 03-JUN-95.
CC PDB; 1EXH; 03-JUN-95.
CC PDB; 1EXP; 27-JAN-97.
CC PDB; 2XYL; 18-MAR-98.
CC PDB; 2HIS; 14-OCT-98.
CC PROSITE; PS00561; CBD_BACTERIAL; 1.
CC PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
CC PFAM; PF00331; glycosyl_hydro3; 1.
CC PFAM; PF00553; CBD.1; 1.
CC CELLULOSE DEGRADATION; HYDROLASE; GLYCOSIDASE; REPEAT; SIGNAL;
CC XYLAN DEGRADATION; 3D-STRUCTURE.
CC
CC SIGNAL 1 41
CC CHAIN 42 484
CC DOMAIN 42 356
CC LINKER ("HINGE") (PRO-THR BOX).
CC DOMAIN 357 376
CC DOMAIN 377 484
CC ACT_SITE 168 168
CC ACT_SITE 274 274

FT DISULFID 208 240
FT DISULFID 302 308
FT DISULFID 382 481
FT MUTAGEN 168 168
SQ SEQUENCE 484 AA; 51291 MW; 2624BI93 CRC32; E->A,D,G: REDUCED ACTIVITY.
Query Match 63.4%; Score 52; DB 1; Length 484;
Best Local Similarity 50.0%; Pred. No. 3.94e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 186 LONGYIETAP 195
Qy 8 FGQGYVQTPF 17
RESULT 9
ID DC11_DROME STANDARD; PRT; 1123 AA.
AC P18169;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE DEFECTIVE CHORION-1 FC125 PROTEIN PRECURSOR.
GN DEC-1.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91032553.
RA WARING G.L., HAWLEY R.J., SCHOENFELD T.;
RT "Multiple proteins are produced from the dec-1 eggshell gene in
RT Drosophila by alternative RNA splicing and proteolytic cleavage
RT events";
RL DEV. BIOL. 142:1-12(1990).
CC -!- FUNCTION: THE DEC-1 FEMALE-STERILE LOCUS PROTEIN IS REQUIRED
CC FOR PROPER ASSEMBLY OF THE DROSOPHILA EGG SHELL.
CC -!- ALTERNATIVE PRODUCTS: MULTIPLE PROTEINS ARE PRODUCED FROM THE
CC DEC-1 GENE BY ALTERNATIVE RNA SPLICING AND PROTEOLYTIC CLEAVAGE
CC EVENTS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M35887; G157182; -
CC PIR; A44766; A44766.
CC FLYBASE; FBgn0000427; dec-1.
CC KW CHORION; EGG SHELL; REPEAT; ALTERNATIVE SPLICING; SIGNAL.
CC SIGNAL 1 23
CC CHAIN 24 1123 DEFECTIVE CHORION-1 FC125 PROTEIN.
CC DOMAIN 493 788 12 X 26 AA TANDEM REPEATS.
CC SEQUENCE 1123 AA; 127959 MW; 99218239 CRC32;
Query Match 63.4%; Score 52; DB 1; Length 1123;
Best Local Similarity 50.0%; Pred. No. 3.94e+00;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 1098 FIQGYAESPY 1107
Qy 8 FGQGYVQTPF 17
RESULT 10
ID RL3_DROME STANDARD; PRT; 415 AA.
AC O16797;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE 60S RIBOSOMAL PROTEIN L3.
GN RPL3.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RA CHAN H.Y.E., ZHANG Y., HOHEISEL J.D., O'KANE C.J.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -- FUNCTION: THE L3 PROTEIN IS A COMPONENT OF THE LARGE SUBUNIT OF
CC CYTOPLASMIC RIBOSOMES.
CC -- SIMILARITY: BELONGS TO THE L3P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF016835; G2384754; -
DR FLYBASE: FBgn0020910; RPL3.
DR PROSITE: PS00474; RIBOSOMAL_L3; 1.
DR PFAM: PF00297; L3; 1.
KW RIBOSOMAL PROTEIN.
FT INIT_MET 0 BY SIMILARITY.
SQ SEQUENCE 415 AA; 46784 MW; FAD30713 CRC32;

Query Match 62.2%; Score 51; DB 1; Length 415;
Best Local Similarity 71.4%; Pred. No. 6.29e+00;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 90 GYIETPF 96
QY 11 GYVQTPF 17

RESULT 11
ID YCR6_YEAST STANDARD; PRT: 743 AA.
AC P25353;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DE 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 84.9 KD PROTEIN IN PMP1-PEN2 INTERGENIC REGION.
GN YCR026C OR YCR26C OR YCR246.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE OF 1-244 FROM N.A.
RA POHL F., RICHTERICH P., WURST H.;
RL SUBMITTED (MAR-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE OF 244-743 FROM N.A.
RX MEDLINE: 92245758.
RA BOLLE P.-A., GILLIQUET V., BERBEN G., DUMONT J., HILGER F.;
RT "The complete sequence of K3B, a 7.9 kb fragment between PGK1 and
RT CRY1 on chromosome III, reveals the presence of seven open reading
RT frames.";
RL YEAST 8:205-213(1992).
CC -- SIMILARITY: TO YEAST YEL016C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: X59720; E264490; -
DR PIR: S19437; S19437.
DR PIR: S27380; S27380.
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 114 135 POTENTIAL.
SQ SEQUENCE 743 AA; 84915 MW; 957CE7AF CRC32;

Query Match 62.2%; Score 51; DB 1; Length 743;
Best Local Similarity 60.0%; Pred. No. 6.29e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Db 563 FPGYIE-PF 571
QY 8 FCGYVQTPF 17

RESULT 12
ID RPOD_EUGGR STANDARD; PRT: 830 AA.
AC P23581;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE DNA-DIRECTED RNA POLYMERASE BETA* CHAIN (EC 2.7.7.6).
GN RPOC2.
OS EUGLENA GRACILIS.
OC CHLOROPLAST.
OC EUKARYOTA; EUGLENOZOA; EUGLENIDA; EUGLENALES; EUGLENA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Z;
RX MEDLINE: 90245579.
RA YEPIZ-PLASCENCIA G.M., RADEBAUGH C.A., HALLICK R.B.;
RT "The Euglena gracilis chloroplast rpoB gene. Novel gene organization
RT and transcription of the RNA polymerase subunit operon.";
RL NUCLEIC ACIDS RES. 18:1869-1878(1990).
RN [2]
RP SEQUENCE OF 586-830 FROM N.A.
RC STRAIN-Z;
RX MEDLINE: 82142472.
RA OROZCO E.M., HALLICK R.B.;
RT "Euglena gracilis chloroplast transfer RNA transcription units. II.
RT Nucleotide sequence analysis of a tRNAVal-tRNAAsn-tRNAArg-tRNALeu
RT gene cluster.";
RL J. BIOL. CHEM. 257:3265-3275(1982).
CC -- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +
CC RNA(N).
CC -- SUBUNIT: IN CHLOROPLAST THE RNA POLYMERASE IS COMPOSED OF FOUR
CC SUBUNITS: ALPHA, BETA, BETA', AND BETA".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X17191; G311711; -
DR EMBL: M22010; G336887; -
DR EMBL: X70810; G415792; -
DR PIR: S19259; RNEGB2.
DR PIR: S34554; S34554.
DR PFAM: PF00623; RNA_pol_A; 1.
DR MENDEL: 5016; EUGgr:RPOC2;1.
KW TRANSCRIPTION; DNA-DIRECTED RNA POLYMERASE; CHLOROPLAST.
FT CONFLICT 593 I -> II (IN REF. 2).
FT CONFLICT 688 L -> R (IN REF. 2).
SQ SEQUENCE 830 AA; 94756 MW; 003F633F CRC32;

Query Match 62.2%; Score 51; DB 1; Length 830;

PFAM; PF00324; aa_permeases; 1.
TRANSPORT; AMINO-ACID TRANSPORT; TRANSMEMBRANE; INNER MEMBRANE.

FT	INIT_MET	0	0	CYTOPLASMIC.
FT	DOMAIN	1	20	
FT	TRANSMEM	21	41	
FT	DOMAIN	42	54	PERIPLASMIC.
FT	TRANSMEM	55	75	
FT	DOMAIN	76	103	CYTOPLASMIC.
FT	TRANSMEM	104	124	
FT	DOMAIN	125	129	PERIPLASMIC.
FT	TRANSMEM	130	148	
FT	DOMAIN	149	159	CYTOPLASMIC.
FT	TRANSMEM	160	180	
FT	DOMAIN	181	198	PERIPLASMIC.
FT	TRANSMEM	199	217	
FT	DOMAIN	218	243	CYTOPLASMIC.
FT	TRANSMEM	244	263	
FT	DOMAIN	264	282	PERIPLASMIC.
FT	TRANSMEM	283	303	
FT	DOMAIN	304	346	CYTOPLASMIC.
FT	TRANSMEM	347	364	
FT	DOMAIN	365	370	PERIPLASMIC.
FT	TRANSMEM	371	391	
FT	DOMAIN	392	414	CYTOPLASMIC.
FT	TRANSMEM	415	432	
FT	DOMAIN	433	448	PERIPLASMIC.
FT	TRANSMEM	449	469	
FT	DOMAIN	470	488	CYTOPLASMIC.
FT	CONFLICT	121	121	N -> S (IN REF. 1).
SEQ	SEQUENCE	488 AA;	53471 MW;	B0E7230B CRC32;

Query Match 61.0%; Score 50; DB 1; Length 488;
Best Local Similarity 50.0%; Pred. No. 9.95e+00;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 84 YGQNYVEEGF 93
: : : : :
Qy 8 FGQGYVQTGF 17

RESULT 14
ID TRIO_HUMAN STANDARD; PRT; 115 AA.
AC Q15642;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE THYROID RECEPTOR INTERACTING PROTEIN 10 (TRIP10) (FRAGMENT).
OS TRIPIO.
GN HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95295737.
RA LEE J.W., CHOI H.-S., GYURIS J., BRENT R., MOORE D.D.;
RT "Two classes of proteins dependent on either the presence or absence
RT of thyroid hormone for interaction with the thyroid hormone
RT receptor";
RL MOL. ENDOCRINOL. 9:243-254(1995).
CC -1- FUNCTION: THYROID RECEPTOR INTERACTING PROTEINS (TRIPS)
CC SPECIFICALLY INTERACT WITH THE LIGAND BINDING DOMAIN OF THE
CC THYROID RECEPTOR (TR). TRIPIO REQUIRES THE PRESENCE OF THYROID
CC HORMONE FOR ITS INTERACTION.
CC -1- TISSUE SPECIFICITY: EXPRESSED MOST ABUNDANTLY IN SKELETAL MUSCLE,
CC HEART AND LUNG.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement>)

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L40379; G695376; 1.
DR PROSITE; PS00002; SH3; 1.
DR PFAM; PF00018; SH3; 1.
DR SH3 DOMAIN.
FT NON_TER 1 1
FT DOMAIN 60 115 SH3.
SQ SEQUENCE 115 AA; 12848 MW; E678DBE7 CRC32;

Query Match 59.8%; Score 49; DB 1; Length 115;
Best Local Similarity 55.6%; Pred.No. 1.56e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 101 GEGYVPTSY 109
   1:1111:
QY 9 GGGYVQTFF 17

Search completed: Wed Sep 1 16:21:06 1999
Job time : 13 secs.
```

```
FT ACT_SITE 233 233 BY SIMILARITY.
FT CONFLICT 36 37 QA -> RP (IN REF. 2).
SQ SEQUENCE 381 AA; 41835 MW; 6A388BC8 CRC32;

Query Match 59.8%; Score 49; DB 1; Length 381;
Best Local Similarity 62.5%; Pred.No. 1.56e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 296 YGQDYVHT 303
   :11111:
QY 8 FGQGYVOT 15

Search completed: Wed Sep 1 16:21:06 1999
Job time : 13 secs.
```

```
RESULT 15
ID DHB2_MOUSE STANDARD; PRT; 381 AA.
AC P51658; O08898;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ESTRADIOL 17 BETA-DEHYDROGENASE 2 (EC 1.1.1.62) (17-BETA-HSD 2)
DE (17-BETA-HYDROXYSTEROID DEHYDROGENASE 2).
GN HSD17B2 OR EDH17B2.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=LIVER;
RX MEDLINE; 97344259.
RA MUSTONEN M.;
RT *Cloning of mouse 17beta-hydroxysteroid dehydrogenase type 2, and
RT analysing expression of the mRNAs for types 1, 2, 3, 4 and 5 in mouse
RT embryos and adult tissues.;
RL BIOCHEM. J. 325:199-205(1997).
RN [2]
RP SEQUENCE OF 1-358 FROM N.A.
RC STRAIN-BALB/C;
RA STOFFEL W., WEISS B.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: CAPABLE OF CATALYZING THE INTERCONVERSION OF
CC TESTOSTERONE AND ANDROSTENEDIONE, AS WELL AS ESTRADIOL AND
CC ESTRONE. ALSO HAS 20-ALPHA-HSD ACTIVITY. USES NADH WHILE EDH17B3
CC USES NADPH (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ESTRADIOL-17-BETA + NAD(+) -> ESTRONE +
CC NADH.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y09517; E289833; -
DR EMBL; X95685; E223346; -
DR MGD; MGI:1096386; HSD17B2.
DR PROSITE; PS00061; ADH_SHORT; 1.
DR PFAM; PF00106; adh_short; 1.
DR HSSP; P14061; 1FDV.
DR STEROID BIOSYNTHESIS; OXIDOREDUCTASE; NAD; MULTIGENE FAMILY;
KW TRANSMEMBRANE; SIGNAL-ANCHOR.
FT TRANSMEM 4 24 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT NP_BIND 83 112 NAD (BY SIMILARITY).
```

 W A P S N L A
 (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed Sep 1 16:21:25 1999; MasPar time 5.97 Seconds
 91.381 Million cell updates/sec
 Tabular output not generated.

Title: >PCT-US99-13024-2
 Description: (8-17) from PCTUS9913024.pep (8 of 12)
 Perfect Score: 82
 Sequence: 1 FCQGVQTPF 10

Scoring table: PAM 150
 Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: spiremb19

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
 5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
 9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
 13:sp-vertebrate 14:sp-virus

Statistics: Mean 23.872; Variance 31.947; scale 0.747

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	55	67.1	509	1	LONG-CHAIN-FATTY-ACID-PUTATIVE G2-SPECIFIC P	2.79e+00
2	54	65.9	722	3	PUTATIVE G2-SPECIFIC P	4.38e+00
3	53	64.6	209	5	F53f8.3 PROTEIN.	6.83e+00
4	53	64.6	344	2	OUTER MEMBRANE PROTEIN	6.83e+00
5	53	64.6	357	2	MAJOR OUTER MEMBRANE P	6.83e+00
6	53	64.6	357	2	OUTER MEMBRANE PROTEIN	6.83e+00
7	53	64.6	364	2	MAJOR OUTER MEMBRANE P	6.83e+00
8	53	64.6	364	2	MAJOR OUTER MEMBRANE P	6.83e+00
9	53	64.6	365	2	MAJOR OUTER MEMBRANE P	6.83e+00
10	53	64.6	365	2	MAJOR OUTER MEMBRANE P	6.83e+00
11	53	64.6	368	2	MAJOR OUTER MEMBRANE P	6.83e+00
12	53	64.6	368	2	MAJOR OUTER MEMBRANE P	6.83e+00
13	53	64.6	390	2	OUTER MEMBRANE PROTEIN	6.83e+00
14	53	64.6	584	5	F21C3.4 PROTEIN.	6.83e+00
15	53	64.6	686	4	K1AA0347 PROTEIN.	6.83e+00
16	53	64.6	936	2	CYTOTOCHROME C BIOGENESI	6.83e+00
17	52	63.4	60	9	OREF50.	1.06e+01
18	52	63.4	485	2	BETA-1,4-XYLANASE PREC	1.06e+01
19	51	62.2	65	7	LEPIDODACTYLUS MOESTUS	1.63e+01
20	51	62.2	340	2	HYPOTHETICAL 37.1 KD P	1.63e+01

21	51	62.2	400	4	Q92734	TGF PROTEIN.	1.63e+01
22	51	62.2	2301	10	O80784	F13p17.19 PROTEIN.	1.63e+01
23	51	62.2	5060	2	O52545	POLYKETIDE SYNTHASE.	1.63e+01
24	51	62.2	5069	2	O52789	RIFAMYCIN POLYKETIDE S	1.63e+01
25	50	61.0	234	2	O25327	MEMBRANE FUSION PROTEI	2.51e+01
26	50	61.0	241	2	O68961	PUTATIVE MEMBRANE FUSI	2.51e+01
27	50	61.0	566	2	P71093	NEOPULLULANASE.	2.51e+01
28	50	61.0	612	5	O26441	RK2-GLIAL-SPECIFIC HOM	2.51e+01
29	50	61.0	612	5	O24477	REVERSED POLARITY.	2.51e+01
30	50	61.0	613	2	O59933	LIPASE (EC 3.1.1.3).	2.51e+01
31	50	61.0	613	2	O59933	EXTRACELLULAR LIPASE (2.51e+01
32	50	61.0	792	10	O48572	SIMILAR TO MAIZE TRANS	2.51e+01
33	50	61.0	793	5	O22047	TO1B7.6 PROTEIN.	2.51e+01
34	50	61.0	803	5	O46129	TRP-LIKE PROTEIN.	2.51e+01
35	49	59.8	107	1	P94122	HYPOTHETICAL 12.0 KD P	3.82e+01
36	49	59.8	248	2	P75004	REP PROTEIN.	3.82e+01
37	49	59.8	255	2	O50462	U0002M.	3.82e+01
38	49	59.8	347	2	O70080	OMF1 PROTEIN (FRAGMENT	3.82e+01
39	49	59.8	347	2	O69304	OMF1 (FRAGMENT).	3.82e+01
40	49	59.8	360	2	O68623	MAJOR OUTER MEMBRANE P	3.82e+01
41	49	59.8	418	2	P95139	HYPOTHETICAL 45.1 KD P	3.82e+01
42	49	59.8	425	1	O59244	425AA LONG HYPOTHETICA	3.82e+01
43	49	59.8	433	10	O64988	ACETYL COA: BENZYLALCO	3.82e+01
44	49	59.8	545	4	O15184	CDC42-INTERACTING PROT	3.82e+01
45	49	59.8	547	11	P97531	SALT-TOLERANT PROTEIN.	3.82e+01

ALIGNMENTS

RESULT	ID	Q28347	PRELIMINARY:	PRT:	509 AA.
AC	O28347				
DT	01-JAN-1998	(TREMBLREL. 05, CREATED)			
DT	01-JAN-1998	(TREMBLREL. 05, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	LONG-CHAIN-FATTY-ACID--COA LIGASE (FADD-8).				
GN	AF1932.				
OS	ARCHAEOGLOBUS FULGIDUS.				
OC	ARCHAEA; EURYARCHAEOTA; ARCHAEOGLOBALES; ARCHAEOGLOBACEAE;				
OC	ARCHAEOGLOBUS.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-VC-16 / DSM 4304 / ATCC 49558;				
RX	MEDLINE; 98049343.				
RA	KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,				
RA	KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,				
RA	RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIES N.C.,				
RA	FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,				
RA	KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,				
RA	PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,				
RA	OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., MCDONALD L., UTTERBACK T.,				
RA	COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,				
RA	SADON P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,				
RA	MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,				
RA	VENTER J.C.;				
RT	*The complete genome sequence of the hyperthermophilic,				
RT	sulphate-reducing archaeon Archaeoglobus fulgidus.*;				
RL	NATURE 390:364-370(1997).				
DR	EMBL; AB000970; G2648612;				
DR	TIGR; AF1932;				
DR	PFAM; PF00501; AMP-binding; 1.				
KW	HYPOTHETICAL PROTEIN; LIGASE.				
SQ	SEQUENCE 509 AA; 57035 MW; 6CD28ACD CRC32;				

Query Match 67.1%; Score 55; DB 1; Length 509;
 Best Local Similarity 60.0%; Pred. No. 2.79e+00;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 325 FTGYYGVQPF 334

QY 8 FCQGVQTPF 17

```
RESULT 2
ID O13839 PRELIMINARY; PRT: 722 AA.
AC O13839:
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PUTATIVE G2-SPECIFIC PROTEIN KINASE C19E9.02 (EC 2.7.1.-).
GN SPAC19E9.02.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972:
RA BADCOCK K., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: PROTEIN KINASE THAT PLAYS AN IMPORTANT ROLE IN MITOTIC
CC REGULATION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF
CC SER/THR-PROTEIN KINASES. BELONGS TO THE NIMA SUBFAMILY.
DR EMBL: Z98975; E339915; -.
DR PFAM: PF00069; pkinase.1.
KW HYPOTHETICAL PROTEIN; TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE;
KW ATP-BINDING; MITOSIS; NUCLEAR PROTEIN; CELL CYCLE; CELL DIVISION.
FT DOMAIN 4 281
FT NP_BIND 10 18 ATP (BY SIMILARITY).
FT BINDING 33 33 ATP (BY SIMILARITY).
FT ACT_SITE 151 151 BY SIMILARITY.
FT ACT_SITE 722 AA; 82689 MW; 286A8C12 CRC32;
SQ SEQUENCE 722 AA; 82689 MW; 286A8C12 CRC32;

Query Match 65.9%; Score 54; DB 3; Length 722;
Best Local Similarity 60.0%; Pred. No. 4.38e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 182 FTQSYVGTPY 191
QY 8 FGQGYVQTPF 17
| | | | | |
| | | | | |

RESULT 3
ID O62260 PRELIMINARY; PRT: 209 AA.
AC O62260:
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE F53F8.3 PROTEIN.
GN F53F8.3.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA MATHWENS L.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN K., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38 (1994).
CC -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.

EMBL: Z81547; E1347158; -.
DR PROSITE: PS00518; ZINC_FINGER_C3HC4; 1.
KW ZINC-FINGER.
SQ SEQUENCE 209 AA; 24425 MW; D1978DC1 CRC32;

Query Match 64.6%; Score 53; DB 5; Length 209;
Best Local Similarity 66.7%; Pred. No. 6.83e+00;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 177 GGGYIKCPF 185
QY 9 GGGYVQTPF 17
| | | | | |
| | | | | |

RESULT 4
ID Q48025 PRELIMINARY; PRT: 344 AA.
AC Q48025:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE OUTER MEMBRANE PROTEIN P2 (FRAGMENT).
GN OMEP2.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-12:
RX MEDLINE: 94245383.
RA BELL J., GRASS S., JEANTEUR D., MUNSON R.S.;
RT "Diversity of the P2 protein among nontypeable Haemophilus influenzae
RT isolates.";
RL INFECT. IMMUN. 62:2639-2643 (1994).
DR EMBL: U08206; G559998; -.
FT NON_TER 1 1
FT NON_TER 344 AA; 37781 MW; 257B8DC5 CRC32;
SQ SEQUENCE 344 AA; 37781 MW; 257B8DC5 CRC32;

Query Match 64.6%; Score 53; DB 2; Length 344;
Best Local Similarity 50.0%; Pred. No. 6.83e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 58 YAGGYLETRF 67
QY 8 FGQGYVQTPF 17
| | | | | |
| | | | | |

RESULT 5
ID Q48225 PRELIMINARY; PRT: 357 AA.
AC Q48225:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MAJOR OUTER MEMBRANE PROTEIN P2 (FRAGMENT).
GN P2.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-3221B:
RX MEDLINE: 94018553.
RA DUIM B., DANKERT J., JANSEN H., VAN ALPHEN L.;
RT "Genetic analysis of the diversity in outer membrane protein P2 of
RT non-encapsulated Haemophilus influenzae.";
RL MICROB. PATHOG. 14:451-462 (1993).
DR EMBL: X73392; G860954; -.
FT NON_TER 357 357
FT NON_TER 357 357
SQ SEQUENCE 357 AA; 39093 MW; 95AD9567 CRC32;

Query Match 64.6%; Score 53; DB 2; Length 357;
Best Local Similarity 50.0%; Pred. No. 6.83e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

Db 78 YAQGYLETRF 87
::|::|::|
QY 8 FGQGYVQTF 17

RESULT 6 PRELIMINARY; PRT; 357 AA.
ID O01452;
AC O01452;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR.
GN OMP2
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93084371.
RA SIKEMA D.J., MURPHY T.F.;
RT "Molecular analysis of the P2 porin protein of nontypeable
RT Haemophilus influenzae.";
RL INFECT. IMMUN. 60:5204-5211(1992).
DR EMBL: M93270; G148967; -;
KW OUTER MEMBRANE; PORIN; SIGNAL.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 357 OUTER MEMBRANE PROTEIN P2.
SQ SEQUENCE 357 AA; 39215 MW; C5E6B40D CRC32;

Query Match 64.6%; Score 53; DB 2; Length 357;
Best Local Similarity 50.0%; Pred. No. 6.83e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 YAQGYLETRF 87
::|::|::|
QY 8 FGQGYVQTF 17

RESULT 7 PRELIMINARY; PRT; 364 AA.
ID O68632;
AC O68632;
DT 01-AUG-1998 (TREMREL. 07, CREATED)
DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE MAJOR OUTER MEMBRANE PORIN P2 VARIANT (FRAGMENT).
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94293786.
RA DUIM B., VAN ALPHEN L., EIJK P., JANSSEN H.M., DANKERT J.;
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major
RT outer membrane protein P2 in patients with chronic bronchitis is
RT caused by point mutations.";
RL MOL. MICROBIOL. 11:1181-1189(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-30-77C;
RA REGELINK A.G., DAHAN D., MOLLER L.V.M., COULTON J.W., EIJK P.,
RA VAN ULSEN P., DANKERT J., VAN ALPHEN L.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF052550; G2981123; -;
FT NON_TER 1 1
FT NON_TER 364 364
SQ SEQUENCE 364 AA; 40097 MW; 2D97185B CRC32;

Query Match 64.6%; Score 53; DB 2; Length 364;
Best Local Similarity 50.0%; Pred. No. 6.83e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 71 YAQGYLETRF 80

QY 8 FGQGYVQTF 17
::|::|::|

RESULT 8 PRELIMINARY; PRT; 364 AA.
ID O68634;
AC O68634;
DT 01-AUG-1998 (TREMREL. 07, CREATED)
DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE MAJOR OUTER MEMBRANE PORIN P2 VARIANT (FRAGMENT).
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-32-82E;
RX MEDLINE: 94293786.
RA DUIM B., VAN ALPHEN L., EIJK P., JANSSEN H.M., DANKERT J.;
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major
RT outer membrane protein P2 in patients with chronic bronchitis is
RT caused by point mutations.";
RL MOL. MICROBIOL. 11:1181-1189(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-32-82E;
RA REGELINK A.G., DAHAN D., MOLLER L.V.M., COULTON J.W., EIJK P.,
RA VAN ULSEN P., DANKERT J., VAN ALPHEN L.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF052552; G2981127; -;
FT NON_TER 1 1
FT NON_TER 364 364
SQ SEQUENCE 364 AA; 39839 MW; 147EB78E CRC32;

Query Match 64.6%; Score 53; DB 2; Length 364;
Best Local Similarity 50.0%; Pred. No. 6.83e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 71 YAQGYLETRF 80
::|::|::|
QY 8 FGQGYVQTF 17

RESULT 9 PRELIMINARY; PRT; 365 AA.
ID O68635;
AC O68635;
DT 01-AUG-1998 (TREMREL. 07, CREATED)
DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE MAJOR OUTER MEMBRANE PORIN P2 VARIANT (FRAGMENT).
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-32-82F;
RX MEDLINE: 94293786.
RA DUIM B., VAN ALPHEN L., EIJK P., JANSSEN H.M., DANKERT J.;
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major
RT outer membrane protein P2 in patients with chronic bronchitis is
RT caused by point mutations.";
RL MOL. MICROBIOL. 11:1181-1189(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-32-82F;
RA REGELINK A.G., DAHAN D., MOLLER L.V.M., COULTON J.W., EIJK P.,
RA VAN ULSEN P., DANKERT J., VAN ALPHEN L.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF052553; G2981129; -;
FT NON_TER 1 1
FT NON_TER 365 365
SQ SEQUENCE 365 AA; 39953 MW; 3567EC05 CRC32;

Query Match 64.6%; Score 53; DB 2; Length 364;
Best Local Similarity 50.0%; Pred. No. 6.83e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Query Match 64.6%; Score 53; DB 2; Length 365;
 Best Local Similarity 50.0%; Pred. No. 6.83e+00;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 71 YAQGYLETRF 80
 QY 8 FGQGYVQTPF 17

RESULT 10
 ID O68631 PRELIMINARY; PRT; 365 AA.
 AC O68631; 1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE MAJOR OUTER MEMBRANE PORIN P2 VARIANT (FRAGMENT).
 OS HAEMOPHILUS INFLUENZAE.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
 OC HAEMOPHILUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=30-77B;
 RX MEDLINE; 94293786.
 RA DUIM B., VAN ALPHEN L., EIJK P., JANSEN H.M., DANKERT J.;
 RT "Antigenic drift of non-encapsulated Haemophilus influenzae major
 outer membrane protein P2 in patients with chronic bronchitis is
 caused by point mutations.";
 RL MOL. MICROBIOL. 11:1181-1189(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=30-77B;
 RA REGELINK A.G., DAHAN D., MOLLER L.V.M., COULTON J.W., EIJK P.,
 RA VAN ULSEN P., DANKERT J., VAN ALPHEN L.;
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AF052549; G2981121; -;
 FT NON_TER 1 365
 FT NON_TER 365 365
 SQ SEQUENCE 365 AA; 40098 MW; 4ECA46B6 CRC32;

Query Match 64.6%; Score 53; DB 2; Length 365;
 Best Local Similarity 50.0%; Pred. No. 6.83e+00;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 71 YAQGYLETRF 80
 QY 8 FGQGYVQTPF 17

RESULT 11
 ID O68633 PRELIMINARY; PRT; 368 AA.
 AC O68633; 1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE MAJOR OUTER MEMBRANE PORIN P2 VARIANT (FRAGMENT).
 OS HAEMOPHILUS INFLUENZAE.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
 OC HAEMOPHILUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=30-77F;
 RX MEDLINE; 94293786.
 RA DUIM B., VAN ALPHEN L., EIJK P., JANSEN H.M., DANKERT J.;
 RT "Antigenic drift of non-encapsulated Haemophilus influenzae major
 outer membrane protein P2 in patients with chronic bronchitis is
 caused by point mutations.";
 RL MOL. MICROBIOL. 11:1181-1189(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=30-77F;
 RA REGELINK A.G., DAHAN D., MOLLER L.V.M., COULTON J.W., EIJK P.,
 RA VAN ULSEN P., DANKERT J., VAN ALPHEN L.;
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; AF052551; G2981125; -;
 FT NON_TER 1 368
 FT NON_TER 368 368
 SQ SEQUENCE 368 AA; 40510 MW; 37391D02 CRC32;

Query Match 64.6%; Score 53; DB 2; Length 368;
 Best Local Similarity 50.0%; Pred. No. 6.83e+00;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 71 YAQGYLETRF 80
 QY 8 FGQGYVQTPF 17

RESULT 12
 ID O68630 PRELIMINARY; PRT; 368 AA.
 AC O68630; 1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE MAJOR OUTER MEMBRANE PORIN P2 VARIANT (FRAGMENT).
 OS HAEMOPHILUS INFLUENZAE.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
 OC HAEMOPHILUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=30-77A;
 RX MEDLINE; 94293786.
 RA DUIM B., VAN ALPHEN L., EIJK P., JANSEN H.M., DANKERT J.;
 RT "Antigenic drift of non-encapsulated Haemophilus influenzae major
 outer membrane protein P2 in patients with chronic bronchitis is
 caused by point mutations.";
 RL MOL. MICROBIOL. 11:1181-1189(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=30-77A;
 RA REGELINK A.G., DAHAN D., MOLLER L.V.M., COULTON J.W., EIJK P.,
 RA VAN ULSEN P., DANKERT J., VAN ALPHEN L.;
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AF052548; G2981119; -;
 FT NON_TER 1 368
 FT NON_TER 368 368
 SQ SEQUENCE 368 AA; 40556 MW; 14B756EE CRC32;

Query Match 64.6%; Score 53; DB 2; Length 368;
 Best Local Similarity 50.0%; Pred. No. 6.83e+00;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 71 YAQGYLETRF 80
 QY 8 FGQGYVQTPF 17

RESULT 13
 ID Q01450 PRELIMINARY; PRT; 390 AA.
 AC Q01450; 1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE OUTER MEMBRANE PROTEIN P2 PRECURSOR.
 GN OMPP2.
 OS HAEMOPHILUS INFLUENZAE.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
 OC HAEMOPHILUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93084371.
 RA SIKKEMA D.J., MURPHY T.F.;
 RT "Molecular analysis of the P2 porin protein of nontypeable
 Haemophilus influenzae.";
 RL INFECT. IMMUN. 60:5204-5211(1992).
 DR EMBL; M93268; G148963; -;
 DR PFAM; PF00267; Gram-ve_porins; 1.


```

KW  OUTER MEMBRANE; PORIN; SIGNAL.
FT  SIGNAL 1 20 POTENTIAL.
FT  CHAIN 21 390 OUTER MEMBRANE PROTEIN P2.
SQ  SEQUENCE 390 AA; 42941 MW; 0E424362 CRC32;

Query Match 64.6%; Score 53; DB 2; Length 390;
Best Local Similarity 50.0%; Pred. No. 6.83e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 78 YAOGYLETRF 87
   ::::|:|
QY 8 FGQYVQTPF 17

RESULT 14
ID Q19672 PRELIMINARY; PRT: 584 AA.
AC Q19672;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE F2IC3.4 PROTEIN.
GN F2IC3.4.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA MCMURRAY A.;
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFFEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SHALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
DR EMBL; Z1261; E1345956;
SQ SEQUENCE 584 AA; 66338 MW; 10A646D4 CRC32;

Query Match 64.6%; Score 53; DB 5; Length 584;
Best Local Similarity 87.5%; Pred. No. 6.83e+00;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 243 GNGYVQTP 250
   |:::|:|
QY 9 GQGYVQTP 16

RESULT 15
ID O60294 PRELIMINARY; PRT: 686 AA.
AC O60294;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE KIAA0547 PROTEIN.
GN KIAA0547.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-BRAIN.
RX MEDLINE; 98290545.
RA NAGASE T., ISHIKAWA K., MIYAJIMA N., TANAKA A., KOTANI H., NOMURA N.,

```

```

RA OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA RES. 5:31-39(1998).
DR EMBL; AB011119; D1026403;
SQ SEQUENCE 686 AA; 75583 MW; B5C954CA CRC32;

Query Match 64.6%; Score 53; DB 4; Length 686;
Best Local Similarity 66.7%; Pred. No. 6.83e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 29 ARGVVDPPF 37
   ::::|:|
QY 9 GQGYVQTPF 17

Search completed: Wed Sep 1 16:22:04 1999
Job time : 39 secs.

```

(TM)

Result	Query	Score	Match	Length	DB	ID	Description	Pred. No.
1	73	31.5	272	2	R07522	Alpha-Trichosanthin e	2.30e+01	
2	73	31.5	272	5	R25577	Ribosome inactivating	2.30e+01	
3	73	31.5	345	10	R52824	GTP-cyclohydrolase II	2.30e+01	
4	70	30.2	820	13	R71802	N-acetylchaperosane I	4.14e+01	
5	68	29.3	119	34	W77551	Nickel-binding peripl	6.11e+01	
6	68	29.3	129	37	W72935	Mycobacterium tubercu	6.11e+01	
7	68	29.3	1252	14	R80530	B. sphaericus SLP	7.11e+01	
8	67	28.9	371	29	W41513	N. meningitidis alpha	7.44e+01	
9	66	28.4	248	23	W21704	Luffin-A.	8.99e+01	
10	66	28.4	248	7	R37294	Plant type I RIP Luff	8.99e+01	
11	66	28.4	248	14	R74180	Type I ribosome-inact	8.99e+01	
12	66	28.4	248	3	R12468	Luffa cylindrica bion	8.99e+01	
13	66	28.4	248	26	R63906	Type I ribosome-inact	8.99e+01	
14	66	28.4	248	24	W25141	Luffin-A (a ribosome	8.99e+01	
15	66	28.4	276	10	R53731	Luffin (ribosome inac	8.99e+01	
16	66	28.4	277	6	R29309	Prod. of the luffin-f	8.99e+01	

```
CC amino-2,4-(1H,3H)-pyrimidindion);
CC - rib-4 gene: 6,7-dimethyl-8-ribityllumazine (DMRL)-synthase;
CC - rib-5 gene: riboflavin-synthase;
CC - rib-3 gene: L-3,4-dihydroxy-2-butanone-4-phosphate (DBP)-synthase.
SQ Sequence 345 AA;

Query Match 31.5%; Score 73; DB 10; Length 345;
Best Local Similarity 50.0%; Pred. No. 2.30e+01;
Matches 14; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

Db 114 glale-fddstgellaskattadnhdnt 140
||| ||||| :||| : |||
QY 59 GLHLRVDPSTGALVDSKSYAFSTSDNT 86

RESULT 4
ID R71802 standard; Protein; 820 AA.
AC AC
DT 19-OCT-1995 (first entry)
DE N-acetylheparosan fragmentation enzyme.
KW N-acetylheparosan; polysaccharide K5; fragmentation; degradation; pharmaceutical starting material.
OS Escherichia coli (K5) SEBR 3282 (NCMC I-1013).
FH Key Location/Qualifiers
FT binding_site 384..387
FT /note= "Asn-X1-X2-Ser motif (X1-Ile, Leu or Tyr and X2-Val, Ile or Ala)"
FT binding_site 411..414
FT /note= "Asn-X1-X2-Ser motif (X1-Ile, Leu or Tyr and X2-Val, Ile or Ala)"
FT binding_site 433..436
FT /note= "Asn-X1-X2-Ser motif (X1-Ile, Leu or Tyr and X2-Val, Ile or Ala)"
FT binding_site 461..464
FT /label= potential_lipoprotein_attachment_site
FT /note= "Asn-X1-X2-Ser motif (X1-Ile, Leu or Tyr and X2-Val, Ile or Ala)"
FT binding_site 495..498
FT /note= "Asn-X1-X2-Ser motif (X1-Ile, Leu or Tyr and X2-Val, Ile or Ala)"
FT region 804..809
FT /note= "strongly charged C-terminal region"
FR FR2709132-A.
PN PN
PD 24-FEB-1995.
PF 17-AUG-1993; 010050.
PR 17-AUG-1993; FR-010050.
PA (SNFI ) ELF SANOFI.
PI Legoux R, Leiong P, Salome MLV;
DR WPI: 95-100554/14.
DR N-PSDB; Q86268.
PT PT
PT New DNA encoding enzyme for fragmenting N-acetyl-heparosan - giving products useful as pharmaceutical starting material, also related protein, vector, transformed cells, etc.
PT Claim 2; Page 38-40; 54pp; French.
PS An enzyme which degrades high mol.wt. N-acetylheparosan into fragments of lower mol.wt. (more suitable for use as pharmaceutical starting materials) can be obtained from E.coli (K5) SEBR 3282. The DNA sequence (Q86268) coding for the enzyme was isolated in a plasmid (p838.7) which has been deposited in E.coli RRI strain as CNMC I-1352. The deduced amino acid sequence (R71802) has homology with an exo-poly-alpha-D-galacturonidase from Erwinia chrysanthemi.
SQ Sequence 820 AA;

Query Match 30.2%; Score 70; DB 13; Length 820;
Best Local Similarity 34.3%; Pred. No. 4.14e+01;
Matches 12; Conservative 9; Mismatches 13; Indels 1; Gaps 1;

Db 197 gvsikfagqngilndnk-d-aftkshsfsvfpv 230
|: :: | | | | | | | | | | | | | |
QY 59 GLHLRVDPSTGALVDSKSYAFSTSDNTTSAFVS 93

RESULT 5
```

```

DR WPI: 98-542705/46.
DR N-PSDB: V63945.
DR PT New isolated mycobacteria polypeptides and nucleic acids - used for
DR PT developing products for the diagnosis of or vaccination against
DR PT mycobacterial infections, particularly tuberculosis
PS Claim 1; Page 223; 163pp; English.
CC The present sequence represents a Mycobacterium tuberculosis protein.
CC Products from the present invention, which describes protein fragments
CC and nucleic acid fragments derived from M.tuberculosis, can be used in
CC the detection of and prevention of mycobacterial infections. In
CC particular, the proteins and nucleic acids can be used for the diagnosis
CC of or vaccination against tuberculosis caused by M. tuberculosis,
CC M. africanum or M. bovis.
CC M. africanum or M. bovis.
SQ Sequence 129 AA;

Query Match 29.3%; Score 68; DB 37; Length 129;
Best Local Similarity 26.5%; Pred.No. 6.11e+01;
Matches 9; Conservative 12; Mismatches 12; Indels 1; Gaps 1;

Db 9 ltlrllaasagilaa-afaaagapagpvdafia 41
QY 60 LHLRFDPSTGALVDSKSYAFSTSNDDTSAAFVS 93
      |||: ||| | : || : ||::
      |||: ||| | : || : ||::

RESULT 7
ID R80530 standard; Protein; 1252 AA.
AC R80530;
AD 22-DEC-1995 (first entry)
DE B. sphaericus SLP.
KW Surface layer protein; SLP; fusion protein; vaccine; antigen;
KW surface expression; epitope.
OS Bacillus sphaericus.
FH Key Location/Qualifiers
FT peptide 1..30
FT /label= Sig_peptide
PD W09519371-A2.
PN 20-JUL-1995.
PR 13-JAN-1995; E00147.
PR 14-JAN-1994; GB-000650.
PA (SOLV ) SOLVAY SA.
PI Dablaere RV, Desomer J, Dhaese P;
DR WPI: 95-263827/34.
DR N-PSDB: Q99430.
DR PT Host cell expressing surface layer protein fusion protein - used for
DR PT host presentation of antigens and vaccine prodn.
PS Disclosure; Fig.6; 95pp; English.
CC A probe based on the N-terminal sequence of B. sphaericus P-1 (LMG
CC P-13855) surface layer protein was used to screen an HindIII-
CC generated library to isolate the slp gene. Promoter regions
CC of the gene are used in genetic constructs providing surface
CC expression of heterologous proteins in P-1 hosts.
CC Sequence 1252 AA;

Query Match 29.3%; Score 68; DB 14; Length 1252;
Best Local Similarity 31.3%; Pred.No. 6.11e+01;
Matches 10; Conservative 12; Mismatches 8; Indels 2; Gaps 2;

Db 778 kvliatgiavntdkgdyaf-takeat-afita 807
QY 62 LRWFDPSTGALVDSKSYAFSTSNDDTSAAFVS 93
      ::|::: : |::| |::| |::| :
      ::|::: : |::| |::| |::| :

RESULT 8
ID W41513 standard; Protein; 371 AA.
AC W41513;
AD 17-JUN-1998 (first entry)
DE N. meningitidis alpha-2,3-sialyltransferase protein.
KW Alpha-2,3-sialyltransferase; sialic acid; acceptor; lipid;
KW biologically active oligosaccharide; sialyl-modified protein.
OS Neisseria meningitidis.
PN W09747749-AL.
PD 18-DEC-1997.
PF 10-JUN-1997; CA0390.

```

```
PR 06-JUN-1997: US-872485.
PR 10-JUN-1996: US-019520.
PA (CAN ) NAT RES COUNCIL CANADA.
PI Gilbert M, Jennings MP, Wakarchuk WW, Young NM;
DR WPI: 98-052313/05.
PR N-PSDB: V04125.
PT Nucleic acid sequence encoding Neisseria
PT alpha-2,3-sialyltransferase - useful to add sialic acid to acceptor
PT with terminal galactose residue for synthesis of biologically active
PT oligosaccharide 38; 50pp; English.
PS Claim 31; Page 38; 50pp; English.
CC This sequence represents an alpha-2,3-sialyltransferase which has been
CC isolated from Neisseria meningitidis. The protein can be used as a
CC reagent for adding a sialic acid residue to an acceptor having a
CC terminal galactose residue, e.g. in synthesis of biologically active
CC oligosaccharides or sialyl-modified proteins or lipids. The nucleic
CC acid sequence can be used for the recombinant production of
CC alpha-2,3-sialyltransferase.
SQ Sequence 371 AA;

Query Match 28.9%; Score 67; DB 29; Length 371;
Best Local Similarity 47.1%; Pred. No. 7.42e+01;
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Db 160 iktfdgtgnligssy 176
   ::||::||::||::||
QY 62 LRVFDPSTGALVDSKSY 78

RESULT 9
ID W21704 standard; Protein: 248 AA.
AC W21704;
DT 26-SEP-1997 (first entry)
DE Luffin-A.
KW pro-Ribosome inactivating protein; proRIP; peptide linker; cancer;
KW inactivation; eukaryotic ribosome; alpha fragment; beta fragment;
KW inhibitor; protein synthesis; N-glycosidase; glycosidic bond; liver;
KW rat; ribosomal 28S RNA; cellular proliferation; HIV-infected T cell.
OS Luffa cylindrica.
FH Key
FT region
FT 138..148
FT /note= "Position of possible insertion of internal
FT peptide linker sequence"

US5635384-A.
PN 03-JUN-1997.
PF 11-JUN-1990: US-535636.
PR 11-JUN-1990: US-535636.
PR 09-DEC-1992: US-587927.
PR 26-JAN-1995: US-578761.
PA (DOMC ) DOWELANCO.
PI Hey TD, Morgan AER, Walsh TA;
DR WPI: 97-309831/28.
PT Inactive precursor of maize ribosome-inactivating protein - also
PT chimeric ribosome-inactivating protein precursors containing
PT internal linker sequences
PS Claim 2; Column 117-118; 121pp; English.
CC The sequences given in W21698-710 represent Ribosome Inactivating
CC proteins (RIP's), which may be used in the construction of the
CC proRIP of the invention. The proRIP has a selectively removable,
CC internal peptide linker. The precursor sequence is incapable of
CC inactivating eukaryotic ribosomes, but can be converted by removal
CC of the linker into a protein having alpha and beta fragments and being
CC capable of inactivating eukaryotic ribosomes. RIPs are potent
CC inhibitors of eukaryotic protein synthesis. They possess a highly
CC specific N-glycosidase activity which cleaves the glycosidic bond of
CC adenine 4324 of rat liver ribosomal 28S RNA. RIP's selectively inhibit
CC cellular proliferation of cells, e.g. cancer cells and HIV-infected T
CC cells. The inactive proRIP proteins make it possible to provide protein
CC synthesis inhibitors with uses in practical and improved ways not before
CC possible. The RIP can be used to make cytotoxic conjugates.
SQ Sequence 248 AA;

Query Match 28.4%; Score 66; DB 23; Length 248;
Best Local Similarity 39.1%; Pred. No. 8.99e+01;
Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Db 73 gylvnstsyffnesdaklasgvy 95
   |::||::||::||
QY 70 GALVDSKSYAFSTNDTTSAAVF 92

RESULT 10
ID R37294 standard; protein: 248 AA.
AC R37294;
DT 13-SEP-1993 (first entry)
DE Plant type I RIP Luffin.
KW Type I ribosome-inactivating protein; ricin; gelonin; momordin;
KW immunoconjugate; autoimmune disease; cell killing; toxin; loofah.
OS Luffa.
PN WO9309130-A.
PD 13-MAY-1993.
PF 04-NOV-1992; U09487.
PR 04-NOV-1991; US-787567.
PR 19-JUN-1992; US-901707.
PA (XOMA ) XOMA CORP.
PI Bernhard SL, Better MD, Carroll SF, Lane JA, Lei SP;
DR WPI: 93-167617/20.
PT Analogues of type I ribosome inactivating protein - useful as
PT cytotoxic agents, immuno toxins for treating autoimmune diseases,
PT cancer, graft versus host disease and selective cell killing in-vivo
PS Disclosure; Page 95-96; 163pp; English.
CC The invention covers analogues of Type I RIPS. Luffin is a
CC Type I RIP and the analogues of the invention have a cysteine
CC available for intermolecular disulphide bonding at an amino acid
CC position corresp. to a position not naturally available for bonding;
CC the cyst residue is located in the C-terminal region of the analogue
CC between a position corresp. to amino acid 251 and the C-terminus of
CC ricin A chain. The analogues are pref. joined via a disulphide
CC linkage to a molecule which specifically binds to a target cell, e.g.
CC an antibody fragment.
SQ Sequence 248 AA;

Query Match 28.4%; Score 66; DB 7; Length 248;
Best Local Similarity 39.1%; Pred. No. 8.99e+01;
Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Db 73 gylvnstsyffnesdaklasgvy 95
   |::||::||::||
QY 70 GALVDSKSYAFSTNDTTSAAVF 92

RESULT 11
ID R74180 standard; protein: 248 AA.
AC R74180;
DT 01-JAN-1996 (first entry)
DE Type I ribosome-inactivating protein luffin.
KW Ribosome inactivating protein; RIP; Type I; cytotoxin; immunotoxin.
OS Plant.
FH Key
FT misc_difference 14
FT /label= Location/Qualifiers
FT /label= invariant residue
FT /note= "in ricin A-chain and the Type I RIPS"
FT misc_difference 22
FT /label= see above
FT misc_difference 70
FT /label= see above
FT misc_difference 110
FT /label= see above
FT misc_difference 131
FT /label= see above
FT misc_difference 157..158
FT /label= see above
FT misc_difference 160
FT /label= see above
FT misc_difference 192
FT /label= see above
```

[illegible]

CC maize RIP, Trichosanthin, Ricin A-chain, Abrin-A A-chain and
 CC Saporin. The RIPs can be used in the construction of therapeutic
 CC toxins targeted to specific cells such as tumour cells via the
 CC attachment of a targeting polypeptide, e.g. a monoclonal antibody.
 CC A further use is in HIV therapy (see US4869903). There is interest
 CC in expressing RIP recombinantly in host eukaryotic cells, because of
 CC the capacity to provide correct post-translational processing. However,
 CC RIPs effectively inhibit protein synthesis in eukaryotic cells resulting
 CC in cell death. Since the inactive RIP proteins are not cytotoxic to
 CC eukaryotic cells, they can be recombinantly expressed in such cells and
 CC then converted to active RIP proteins.
 SQ Sequence 248 AA;

Query Match 28.4%; Score 66; DB 24; Length 248;

Best Local Similarity 39.1%; Pred. No. 8.99e+01;

Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Db 73 gylvnstsyffnesdaklasqv 95

QY 70 GALVDSKSYAFSTNDTTSAAFV 92

RESULT 15

ID R53731 standard; Protein: 276 AA.

AC R53731;

DT 09-FEB-1995 (first entry)

DE Luffin (ribosome inactivating protein).

KW Luffin; ribosome inactivating protein; RIP; loofah; inhibition;

OS protein translation; protein synthesis; viral proliferation.

KS Luffa cylindrica.

FH Key Location/Qualifiers

FT peptide 1..20

FT protein /label= signal_peptide

FT 21..276

FT /label= luffin

FT misc_difference 123

FT /note= "Leu corresponds to TCC codon"

PN J06141867-A.

PD 24-MAY-1994.

PF 04-NOV-1992; 295301.

PR 04-NOV-1992; JP-295301.

PA (TORA) TORAY IND INC.

DR WPI; 94-205026/25.

DR N-PSDB; Q66093.

PT DNA encoding luffin - used for mass production of luffin, a

PT ribosome inactivating protein

PS Claim 2; Page 4-5; 5pp; Japanese.

CC The luffin coding sequence was isolated from a cDNA library prepared

CC from mRNA isolated from seeds of Luffa cylindrica. The cDNA will be

CC useful for recombinant production of luffin, a ribosome inactivating

CC protein. The protein inhibits protein synthesis and viral

CC replication.

SQ Sequence 276 AA;

Query Match 28.4%; Score 66; DB 10; Length 276;

Best Local Similarity 39.1%; Pred. No. 8.99e+01;

Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Db 92 gylvnstsyffnesdaklasqv 114

QY 70 GALVDSKSYAFSTNDTTSAAFV 92

Search completed: Wed Sep 1 16:26:18 1999

Job time : 38 secs.


```

REFERENCE S46565
#authors Skala, J.; van Dyck, L.; Purnelle, B.; Goffeau, A.
#journal Yeast (1994) 10(Suppl.A):S13-S24
#title The sequence of an 8.8 kb segment on the left arm of chromosome II from Saccharomyces cerevisiae reveals four new open reading frames including homologs of animal DNA polymerase alpha-primases and bacterial GTP cyclohydrolase II.

#accession S46565
##molecule_type DNA
##residues 1-345 ##label SKA
##cross-references EMBL:X74738; NID:g511140; PID:g511141
##experimental_source strain S288C
#REFERENCE S50971
#authors Saiz, J.; Santos, M.A.; Plaza, M.A.; Revuelta, J.L.
#submission submitted to the EMBL Data Library, February 1993
#description Cloning and sequencing of the RibI gene from Saccharomyces cerevisiae.

#accession S50971
##molecule_type DNA
##residues 1-180,'N',182-345 ##label SAI
##cross-references EMBL:Z21617; NID:g642218; PID:g642219
#COMMENT This enzyme is involved in riboflavin (vitamin B2) biosynthesis, catalyzing the release of formate and inorganic pyrophosphate under formation of 2,5-diamino-6-ribosylamino-4(3H)-pyrimidine 5'-phosphate from GTP.

GENETICS
#gene SGD:RIB1
##cross-references SGD:S0000129; MIPS:YBL033c
#map_position 2L
CLASSIFICATION #superfamily yeast cyclohydrolase II; cyclohydrolase homology
KEYWORDS GTP; hydrolase; riboflavin biosynthesis
FEATURE
139-309 #domain cyclohydrolase homology #label CYCH
SUMMARY #length 345 #molecular-weight 38346 #checksum 7596

Query Match 31.5%; Score 73; DB 2; Length 345;
Best Local Similarity 50.0%; Pred. No. 2.75e+00;
Matches 14; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

Db 114 GLALE-FDDSTGELLASKATTWDAHNDT 140
||| ||| ||| ||| : ||| : |||
QY 59 GLHLRVDPSTGALVDSKSYAFSTSNDT 86
||||| ||| ||| ||| : ||| : |||

RESULT 3
ENTRY H69418 #type complete
TITLE Hypothetical protein AF1353 - Archaeoglobus fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
DATE 05-Dec-1997 #sequence_revision 03-Dec-1997 #text_change
ACCESSIONS H69418
REFERENCE A69250
#authors Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.B.; Kerlavage, A.R.; Graham, D.E.; Kyriades, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Patterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kalne, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
#journal Nature (1997) 390:364-370
#title The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references EMBL:Z21617; NID:g642218; PID:g642219
#accession H69418
##status preliminary; nucleic acid sequence not shown; translation not shown
```

```
Db      144 LRQLDSHNPLADLSPFAFGSGNASVSAP 172
```

```
##experimental_source strain S288C
REFERENCE S72107
#authors Eide, L.G.; Sander, C.; Prydz, H.
#journal Yeast (1996) 12:1085-1090
#title Sequencing and analysis of a 35.4 kb region on the left arm
of chromosome IV from Saccharomyces cerevisiae reveal 23
open reading frames.
#cross-references MUID:97051598
#accession S72126
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-889 #label EIW
##cross-references EMBL:X59566; NID:g1216215; PID:g225538; PID:g1216235
##note the nucleotide sequence was submitted to the EMBL Data
Library, February 1996
GENETICS
#map_position 4R
#note YDR027C
KEYWORDS transmembrane protein
FEATURE
767-783
SUMMARY
#domain transmembrane #status predicted #label TMW
#length 889 #molecular-weight 101519 #checksum 1524
Query Match 31.0%; Score 72; DB 2; Length 889;
Best Local Similarity 31.0%; Pred. No. 3.86e+00;
Matches 9; Conservative 8; Mismatches 12; Indels 0; Gaps 0;
Db 77 RSPDSRTLSVANSRSGFENETHSGSMDF 105
| | | | | : | | | | | : | | | | |
QY 63 RVFDPSTGALVDSKSYAFSTNDTTSAAF 91
| | | | | : | | | | | : | | | | |
RESULT 10
ENTRY H36812 #type complete
TITLE hypothetical protein ORF64 - saimirine herpesvirus 1 (strain
11)
ORGANISM #formal_name saimirine herpesvirus 1
#note host Saimiri sciureus (common squirrel monkey)
DATE 16-Oct-1992 #sequence_revision 16-Oct-1992 #text_change
09-Sep-1997
ACCESSIONS H36812
REFERENCE H36806
#authors Albrecht, J.
#submission Submitted to the EMBL Data Library, January 1992
#description Primary structure of the herpesvirus saimiri genome.
#accession H36812
##molecule_type DNA
##residues 1-2469 #label ALB
##cross-references GB:X64346; NID:g60320; PID:g60385
REFERENCE A37309
#authors Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.;
Biesinger, B.; Newman, C.; Wittmann, S.; Craxton, M.A.;
Coleman, H.; Fleckenstein, B.; Honess, R.W.
#journal J. Virol. (1992) 66:5047-5058
#title Primary structure of the herpesvirus saimiri genome.
#cross-references MUID:92333688
#contents annotation; protein-coding frames
#note neither protein nor nucleotide sequence is given
GENETICS
#gene 64
#summary #length 2469 #molecular-weight 280165 #checksum 4051
Query Match 31.0%; Score 72; DB 2; Length 2469;
Best Local Similarity 38.7%; Pred. No. 3.86e+00;
Matches 12; Conservative 8; Mismatches 10; Indels 1; Gaps 1;
Db 161 IFDPHSCPLVPNSPAHVFSNVDATLEYLS 191
| | | | | : | | | | | : | | | | |
QY 64 VFDPSTGALV-DSKSYAFSTNDTTSAAFVS 93
| | | | | : | | | | | : | | | | |
RESULT 11
ENTRY D69200 #type complete
```

```
TITLE conserved hypothetical protein MTH751 - Methanobacterium
thermoautotrophicum (strain Delta H)
ORGANISM #formal_name Methanobacterium thermoautotrophicum
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
04-Sep-1998
ACCESSIONS D69200
REFERENCE A69000
#authors Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;
Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.;
Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;
Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, P.;
Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.; Caruso,
A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.;
McDougall, S.; Shimer, G.; Goyal, A.; Pietrokovski, S.;
Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
J.; Reeve, J.N.
#journal J. Bacteriol. (1997) 179:7135-7155
#title Complete genome sequence of Methanobacterium
thermoautotrophicum Delta H: functional analysis and
comparative genomics.
#cross-references MUID:98037514
#accession D69200
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-241 #label MTH
##cross-references GB:AE000854; GB:AE000666; NID:g2621839; PID:g2621840
##experimental_source strain Delta H
GENETICS
#gene MTH751
#start_codon GTG
CLASSIFICATION #superfamily hypothetical protein MJ1374
SUMMARY #length 241 #molecular-weight 26734 #checksum 570
Query Match 30.6%; Score 71; DB 2; Length 241;
Best Local Similarity 38.9%; Pred. No. 5.39e+00;
Matches 14; Conservative 8; Mismatches 11; Indels 3; Gaps 2;
Db 24 AGNIIHT-DPGCALV--RSYQFSDPRKLDAYMVS 56
| | | | | : | | | | | : | | | | |
QY 58 AGLHLRVFDPSTGALVDSKSYAFSTNDTTSAAFVS 93
| | | | | : | | | | | : | | | | |
RESULT 12
ENTRY S69889 #type complete
TITLE hemagglutinin precursor - influenza A virus (isolate
SW/Ehime/1/80 [H1N2])
ORGANISM #formal_name influenza A virus
#variety H1N2 isolate sw/Ehime/1/80
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
01-Feb-1999
ACCESSIONS S69889
REFERENCE S69886
#authors Sugita, S.; Yoshioka, Y.; Itamura, S.; Kanegae, Y.; Oguchi,
K.; Gojobori, T.; Nerome, K.; Oya, A.
#journal J. Mol. Evol. (1991) 32:16-23
#title Molecular evolution of hemagglutinin genes of H1N1 swine and
human influenza A viruses.
#cross-references MUID:91186608
#accession S69889
##status nucleic acid sequence not shown
##molecule_type genomic RNA
##residues 1-566 #label SUG
##cross-references EMBL:X57494
GENETICS
#gene HA1; HA2
#classification #superfamily influenza virus hemagglutinin
#keywords glycoprotein; hemagglutinin; transmembrane protein
FEATURE
1-17
18-343
345-566
530-546
#domain signal sequence #status predicted #label SIG\
#product hemagglutinin HA1 #status predicted #label HA1\
#product hemagglutinin HA2 #status predicted #label HA2\
#domain transmembrane #status predicted #label TM1\
```

S25345	#type complete problemable membrane protein YCR089w - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES	hypothetical protein YCR1102
ORGANISM	#formal_name Saccharomyces cerevisiae
DATE	30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 06-Feb-1998
ACCESSIONS	S25345; S19504
REFERENCE	#authors Wilson, C.; Grisanti, P.; Frontali, L. #journal Yeast (1992) 8:569-575 #title The complete sequence of a 6146 bp fragment of Saccharomyces cerevisiae chromosome III contains two new open reading frames.
#cross-references	MUID:92397594
#accession	S25345
#molecule_type	DNA
#residues	1-1609 ##label WIL
#cross-references	GB:X59720; EMBL:S43845; NID:g1907116; PID:e264634; PID:g1907227
REFERENCE	S19504 #authors Frontali, L.; Grisanti, P. #submission submitted to the Protein Sequence Database, March 1992 #accession S19504
#molecule_type	DNA
#residues	1-1609 ##label PRO
#cross-references	EMBL:X59720; NID:g1907116; PID:e264634; PID:g1907227 MIPS:YCR089w
GENETICS	sgd:FIG2
#gene	
#cross-references	SGD:S0000685; MIPS:YCR089w
#map_position	3R
KEYWORDS	transmembrane protein
FEATURE	4-20 1592-1609 ##domain transmembrane #status predicted #label TM1\ ##domain transmembrane #status predicted #label TM2 #length 1609 #molecular-weight 166047 #checksum 8346
SUMMARY	Query Match 30.6%; Score 71; DB 2; Length 1609; Best Local Similarity 34.2%; Pred. No. 5.39e+00; Matches 13; Conservative 12; Mismatches 12; Indels 1; Gaps 1;
Db	145 SSFELPVTAPSSSLPSSTLTFTSVNPSQSWTSFNS 182 ::: ::: ::: ::: :::
Qy	57 SAGHLRVDPSTGALVDSKSYAFSTSDNTTS-AAFVS 93
RESULT	15
ENTRY	S30708
TITLE	#type complete hypothetical 15.8K protein (corA-rarD intergenic region) - Escherichia coli
ORGANISM	#formal_name Escherichia coli
DATE	02-Dec-1993 #sequence_revision 01-Sep-1995 #text_change 14-Nov-1997
ACCESSIONS	S30708; S30745; C65186
REFERENCE	S30660 #authors Daniels, D.L.; Plunkett III, G.; Burland, V.; Blattner, F.R. #journal Science (1992) 257:771-778 #title Analysis of the Escherichia coli genome: DNA sequence of the region from 84.5 to 86.5 minutes.
#cross-references	MUID:92358234
#accession	S30708
#status	preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type	DNA
#residues	1-138 ##label DAN
#cross-references	EMBL:M87049; NID:g836656; PID:g148217
#note	the nucleotide sequence was submitted to the EMBL Data Library, November 1992
REFERENCE	S30742 #authors Olmori, H. #submission submitted to the EMBL Data Library, January 1993 #description Physical map of the corA region of the E.coli chromosome.

```
##accession      S30745
##status         preliminary
##molecule_type DNA
##residues       1-138 ##label OHM
##cross-references EMBL:L02122; NID:g145575; PID:g145579
REFERENCE
#authors         Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
                  Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
                  Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
                  Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
                  Y.
#journal         Science (1997) 277:1453-1462
#title           The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession       C65186
#status         preliminary; nucleic acid sequence not shown;
                  translation not shown
##molecule_type DNA
##residues       1-138 ##label BLAT
##cross-references GB:AE000457; GB:U00096; NID:g2367294; PID:g1790251;
                  UWGP:b3818
##experimental_source strain K-12, substrain MG1655
GENETICS
#gene            yigG
#map_position     86 min
SUMMARY          #length 138 #molecular-weight 15794 #checksum 1118
Query Match      30.2%; Score 70; DB 2; Length 138;
Best Local Similarity 35.0%; Pred. No. 7.50e+00;
Matches          7; Conservative 6; Mismatches 7; Indels 0; Gaps 0;
Db               14 LRIFPTSGKISRRIYFS 33
                ||| |::: | |
QY               62 LRVFDPSTGALVDSKSYAFS 81
Search completed: Wed Sep 1 16:25:22 1999
Job time : 22 secs.
```

 W A P E S R E H

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed Sep 1 16:23:45 1999; MasPar time 3.53 Seconds
 Tabular output not generated. 296.202 Million cell updates/sec

Title: >PCT-US99-13024-2
 Description: (57-93) from PCTUS9913024.ppt (9 of 12)
 Perfect Score: 232
 Sequence: 1 SAGLHLRVDPSTGALVDSKSYAFSTNDTSAFVS 37

Scoring table: PAM 150
 Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot37
 1:swissprot

Statistics: Mean 33.585; Variance 52.447; scale 0.640

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	74	31.9	1222	1	YMH5_CAEEL	7.18e+01
2	73	31.5	345	1	GCH2_YEAST	1.04e+00
3	72	31.0	300	1	NTRY_AZOBR	1.50e+00
4	72	31.0	2469	1	TEGU_HSVSA	1.50e+00
5	71	30.6	1375	1	GTFC_STRMU	2.16e+00
6	71	30.6	1609	1	YIGG_ECOLI	2.16e+00
7	70	30.2	138	1	YIGG_ECOLI	3.10e+00
8	70	30.2	932	1	YMB4_CAEEL	3.10e+00
9	70	30.2	1545	1	IGA3_HAEIN	3.10e+00
10	70	30.2	1802	1	HRP1_YEAST	3.10e+00
11	69	29.7	566	1	FOBI_YEAST	4.42e+00
12	68	29.3	107	1	PRVB_ESOLU	6.29e+00
13	68	29.3	324	1	HS11_SOYBN	8.91e+00
14	67	28.9	153	1	YH03_YEAST	8.91e+00
15	67	28.9	256	1	YH03_YEAST	8.91e+00
16	67	28.9	533	1	YH03_YEAST	8.91e+00
17	67	28.9	639	1	YH03_YEAST	8.91e+00
18	67	28.9	1091	1	YH03_YEAST	8.91e+00
19	67	28.9	2431	1	YH03_YEAST	8.91e+00
20	67	28.9	3063	1	YH03_YEAST	8.91e+00
21	67	28.9	3067	1	YH03_YEAST	8.91e+00
22	66	28.4	108	1	PRVB_MERBI	1.26e+01
23	66	28.4	163	1	HMCS_DICDI	1.26e+01

24	66	28.4	277	1	RIPA_LUFYC	RIBOSOME-INACTIVATING	1.26e+01
25	66	28.4	539	1	TCPD_MOUSE	T-COMPLEX PROTEIN 1, D	1.26e+01
26	66	28.4	539	1	TCPD_HUMAN	T-COMPLEX PROTEIN 1, D	1.26e+01
27	66	28.4	577	1	SYL_HELPY	PROLYL-TRNA SYNTHETASE	1.26e+01
28	66	28.4	616	1	MUTA_STRCM	METHYLMALONYL-COA MUTA	1.26e+01
29	66	28.4	1430	1	GTFC_STRMU	GLUCOSYLTRANSFERASE-S	1.26e+01
30	65	28.0	108	1	PRVB_MERMR	PARVALBUMIN BETA.	1.26e+01
31	65	28.0	173	1	ELHA_APLCA	ATRIAL GLAND PEPTIDE A	1.76e+01
32	65	28.0	234	1	YODR_MVCTU	HYPOTHETICAL 25.4 KD P	1.76e+01
33	65	28.0	300	1	MOVP_AMVMA	MOVEMENT PROTEIN (CELL	1.76e+01
34	65	28.0	300	1	MOVP_AMVLE	MOVEMENT PROTEIN (CELL	1.76e+01
35	65	28.0	345	1	PURA_METJA	ADENYLOSUCCLINATE SYNTH	1.76e+01
36	65	28.0	429	1	PURA_THIFE	ADENYLOSUCCLINATE SYNTH	1.76e+01
37	65	28.0	480	1	TRA3_STAAT	TRANSPORIN FOR TRANS	1.76e+01
38	65	28.0	536	1	TCPD_FUGRU	T-COMPLEX PROTEIN 1, D	1.76e+01
39	65	28.0	554	1	VL2_HP41	MINOR CAPSID PROTEIN L	1.76e+01
40	65	28.0	566	1	HEMA_TAZIN	HEMAGGLUTININ PRECURSO	1.76e+01
41	65	28.0	2710	1	TOXA_CLODI	COLD SHOCK-LIKE PROTEI	2.47e+01
42	64	27.6	68	1	CSPE_ECOLI	ASPARAGINE SYNTHETASE	2.47e+01
43	64	27.6	585	1	ASNI_LOTJA	HYPOTHETICAL 251.0 KD	2.47e+01
44	64	27.6	2167	1	YCS2_YEAST	HYPOTHETICAL 251.0 KD	2.47e+01
45	64	27.6	4466	1	DYHC_TRIGR	DYNEIN BETA CHAIN, CIL	2.47e+01

ALIGNMENTS

RESULT 1
 ID YMH5_CAEEL STANDARD; PRT: 1222 AA.
 AC P34472;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 136.3 KD PROTEIN F58A4.5 IN CHROMOSOME III.
 GN F58A4.5
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA; RHABDITIDA;
 RN RHABDITIDA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 RC SEQUENCE FROM N.A.
 RX STRAIN-BRISTOL N2;
 RX MEDLINE; 94150718.
 RA WILSON R., AINSKOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
 RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
 RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
 RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SINS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
 RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
 RA WATSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
 RA WOLDMAN P.;
 RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RL NATURE 368:32-38(1994).
 CC -1- SIMILARITY: TO REVERSE TRANSCRIPTASE.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; 222179; E1347484;
 DR PIR; S40977; S40977.
 DR WORMPEP; F58A4.5; CE00221.
 DR PFAM; PF00078; Ivt: 1.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 1222 AA; 136273 MW; 2F6802B7 CRC32;

Query Match 31.9%; Score 74; DB 1; Length 1222;

RT organization between HVS and Epstein-Barr virus.";
RL VIREOLOGY 188:296-310(1992).
CC -!- FUNCTION: TEGUMENT PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EHV-1 24, EBV BFLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; X64346; G60385; -
DR EMBL; M86409; G330993; -
DR PIR; H36812; H36812.
SQ SEQUENCE 2469 AA; 280165 MW; 1B72F5F9 CRC32;

Query Match 31.0%; Score 72; DB 1; Length 2469;
Best Local Similarity 38.7%; Pred. No. 1.50e+00;
Matches 12; Conservative 8; Mismatches 10; Indels 1; Gaps 1;

Db 161 IFDPHSCPVPNSPAHVSTSNVNDIAEYLS 191
QY 64 VEDPSIGALV-DSKSYAFSTSNDDTSAAFVS 93

RESULT 5
ID GTFC STRMU STANDARD; PRT; 1375 AA.
AC P13470; P05427;
DT 01-NOV-1988 (REL. 09, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)
DE (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTFC.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GS-5;
RA MEDLINE; 89137980.
RA UEDA S., SHIROZA T., KURAMITSU H.K.;
RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5";
RL GENE 69:101-109(1988).
RN [2]
RP SEQUENCE OF 1-349 FROM N.A.
RC STRAIN-GS-5;
RA MEDLINE; 87308013.
RA SHIROZA T., UEDA S., KURAMITSU H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. BACTERIOL. 169:4263-4270(1987).
CC -!- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -!- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) -
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- DISEASE: DENTAL CARIES.
CC -!- GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED
CC GLUCOSE AND SOME 1,6 LINKAGES). GTF-S SYNTHESIZES WATER-SOLUBLE
CC GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF
CC GLUCANS.
CC
CC -!- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M22054; G153643; -
DR EMBL; M17361; G153641; -
DR PIR; J00345; J00345.
DR PIR; C33135; C33135.
KW TRANSFERASE; GLYCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.
FT SIGNAL 1 34
FT CHAIN 35 1375 GLUCOSYLTRANSFERASE-SI.
FT DOMAIN 35 1050 CATALYTIC (APPROXIMATE).
FT DOMAIN 1126 1375 GLUCAN-BINDING (APPROXIMATE).
FT REPEAT 1126 1375 2,4 A, 1 C AND 1 AC REPEATS.
FT REPEAT 1126 1159 A REPEAT.
FT REPEAT 1159 1200 A REPEAT.
FT REPEAT 1227 1238 C REPEAT.
FT REPEAT 1253 1303 AC REPEAT.
FT REPEAT 1318 1330 A REPEAT (INCOMPLETE).
SQ SEQUENCE 1375 AA; 153022 MW; 3EA3727E CRC32;

Query Match 30.6%; Score 71; DB 1; Length 1375;
Best Local Similarity 41.4%; Pred. No. 2.16e+00;
Matches 12; Conservative 8; Mismatches 7; Indels 2; Gaps 2;

Db 1093 SSVTGLVFDGKGYVYVYSTGNOAKNAFIS 1121
QY 67 PSTGALV-DSKSYAF-STSNDDTTSAAAFVS 93

RESULT 6
ID FIG2 YEAST STANDARD; PRT; 1609 AA.
AC P25653.
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE FACTOR INDUCED GENE 2.
GN FIG2 OR YCR089W OR YCR89W OR YCR1102.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92397594.
RA WILSON C., GRISANTI P., FRONTALI L.;
RT "The complete sequence of a 6146 bp fragment of Saccharomyces
RL cerevisiae chromosome III contains two new open reading frames.";
RL YEAST 8:569-575(1992).
CC -!- FUNCTION: REQUIRED FOR EFFICIENT MATING.
CC -!- INDUCTION: BY MATING PHEROMONES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X59720; E264634; -
DR PIR; S19504; S19504.
DR PIR; S25345; S25345.
DR SGD; L0003312; FIG2.
SQ SEQUENCE 1609 AA; 166049 MW; DE974CE8 CRC32;

Query Match 30.6%; Score 71; DB 1; Length 1609;
Best Local Similarity 34.2%; Pred. No. 2.16e+00;
Matches 13; Conservative 12; Mismatches 12; Indels 1; Gaps 1;

Db 145 SSSFELPVTAPSSSLPSTSLTFTSVNPSQSWTSFNS 182
QY 57 SAGLHLRVDPSTGALVDSKSYAFSTSNDDTTS-AAFVS 93


```

RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOHLDMAN P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.*;
RL NATURE 368:32-38(1994).
CC -|- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -|- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z19155; G6715;
CC PIR: S28275; S28275.
CC WORPEP: F54G8.4; CE05942.
CC PROSITE: P500518; ZINC_FINGER_C3HC4; 1.
CC PFAM: PF00097; zf-C3HC4; 1.
CC HSSP: P29590; 1BOR.
CC HYPOTHETICAL PROTEIN: ZINC-FINGER; DNA-BINDING; NUCLEAR PROTEIN.
CC ZN_FING 43 83 C3HC4-TYPE.
FT SEQUENCE 932 AA; 104438 MW; 45C877B4 CRC32;
SQ
Query Match 30.2%; Score 70; DB 1; Length 932;
Best Local Similarity 34.4%; Pred. No. 3.10e+00;
Matches 11; Conservative 9; Mismatches 10; Indels 2; Gaps 2;

Db 284 TEGLEYLRNFQPDADLFNRK-LRFSAGDDAA 314
QY 57 SAGLH-LRVFDPSTGALVDSKYAFSTSDTT 87

RESULT 9
ID IGA3_HAEIN STANDARD; PRT; 1545 AA.
AC P45385;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE IMMUNOGLOBULIN A1 PROTEASE PRECURSOR (EC 3.4.21.72) (IGAL PROTEASE).
GN IGA.
OS HAEMOPHILUS INFLUENZAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
OC HAEMOPHILUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HK393 / NCTC 8467 / SEROTYPE B;
RX MEDLINE: 92234949.
RA POULSEN K., REINHOLDT J., KILIAN M.;
RT *A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases.*;
RL J. BACTERIOL. 174:2913-2921(1992).
CC -|- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -|- CATALYTIC ACTIVITY: CLEAVAGE OF IMMUNOGLOBULIN A MOLECULES AT
CC CERTAIN PRO-1-XAA BONDS IN THE HINGE REGION. NO SMALL MOLECULE
CC SUBSTRATES ARE KNOWN.
CC -|- SUBCELLULAR LOCATION: SECRETED.
CC -|- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

RA DANIELS D.L., PLUNKETT G. III, BURLAND V.D., BLATTNER F.R.;
RT *Analysis of the Escherichia coli genome: DNA sequence of the region
RT from 84.5 to 86.5 minutes.*;
RL SCIENCE 257:771-778(1992).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L02122; G145579;
CC EMBL: M87049; G148217;
CC EMBL: AE000457; G1790251;
CC PIR: S30745; S30745.
CC PIR: S30708; S30708.
CC ECOGENE: EGI1465; YIGG.
CC HYPOTHETICAL PROTEIN.
SQ SEQUENCE 138 AA; 15794 MW; C92E2E54 CRC32;

Query Match 30.2%; Score 70; DB 1; Length 138;
Best Local Similarity 35.0%; Pred. No. 3.10e+00;
Matches 7; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

Db 14 LRIFITSNGKISRRIYIS 33
QY 62 LRVFDPSTGALVDSKYAFS 81

RESULT 8
ID YMB4_CABEL STANDARD; PRT; 932 AA.
AC Q03601;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 104.4 KD PROTEIN F54G8.4 IN CHROMOSOME III.
GN F54G8.4.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTERA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE OF 1-574 FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE: 94150718.
RA WILSON R., AINSKOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M87490; G148909; -
KW HYDROLASE; SERINE PROTEASE; TRANSMEMBRANE; ZYMOGEN; SIGNAL;
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1012 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1013 1545 HELPER PEPTIDE (POTENTIAL).
FT ACT SITE 292 292 PROBABLE.
SQ SEQUENCE 1345 AA; 170627 MW; 93EE4AD6 CRC32;

Query Match 30.2%; Score 70; DB 1; Length 1545;
Best Local Similarity 26.7%; Pred. No. 3.10e+00;
Matches 8; Conservative 14; Mismatches 6; Indels 2; Gaps 2;

Db 336 KIYEQISAGSLIGSKTIDYSSWSNGKTSIT 365

QY 63 RVFDP-STGALVDSKS-YAFSTNDTTSAA 90

RESULT 10
ID HKR1_YEAST STANDARD; PRT: 1802 AA.
AC P41809;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HANSENULA MRAKII KILLER TOXIN-RESISTANT PROTEIN 1 PRECURSOR.
GN HKR1 OR YDR420W.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YNN 225;
RX MEDLINE; 94156857.
RA KASAHARA S., YAMADA H., MIO T., SHIRATORI Y., MIYAMOTO C.,
RA YABE T., NAKAJIMA T., ICHISHIMA E., FURUICHI Y.;
RT "Cloning of the Saccharomyces cerevisiae gene whose overexpression
RT overcomes the effects of HM-1 killer toxin, which inhibits
RT beta-glucan synthesis.";
RT J. BACTERIOL. 176:1488-1499(1994).
CC -!- FUNCTION: COULD REGULATE BETA-GLUCAN SYNTHESIS. OVEREXPRESSION
CC PROVIDES RESISTANCE TO HM-1 KILLER TOXIN.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE).
CC -!- PTM: COULD BE O-GLYCOSYLATED IN SERINE/THREONINE RICH DOMAIN.
CC -!- SIMILARITY: SOME, TO YEAST MSB2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; S69101; G545660; -
DR SGD; L0000788; HKR1
KW GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1802 HANSENULA MRAKII KILLER TOXIN-RESISTANT
FT PROTEIN 1.
FT TRANSMEM 1486 1506 POTENTIAL.
FT DOMAIN 23 1478 SER/THR-RICH.
FT DOMAIN 453 788 12 X 28 AA TANDEM REPEATS OF S-[AV]-[P]-
FT V-A-V-S-S-T-Y-T-S-P-A-P-A-I-S-S-T-
FT Y-T-S-S-P.
FT REPEAT 453 480 1 (APPROXIMATE).
FT REPEAT 481 508 2.
FT REPEAT 509 536 3.
FT REPEAT 537 564 4.

FT REPEAT 565 592 5.
FT REPEAT 593 620 6.
FT REPEAT 621 648 7.
FT REPEAT 649 676 8.
FT REPEAT 677 704 9.
FT REPEAT 705 732 10.
FT REPEAT 733 760 11.
FT REPEAT 761 788 12.
FT CARBOHYD 24 24 POTENTIAL.
FT CARBOHYD 1252 1252 POTENTIAL.
FT CARBOHYD 1293 1293 POTENTIAL.
FT CARBOHYD 1342 1342 POTENTIAL.
FT CARBOHYD 1400 1400 POTENTIAL.
SQ SEQUENCE 1802 AA; 188890 MW; 84B6C85F CRC32;

Query Match 30.2%; Score 70; DB 1; Length 1802;
Best Local Similarity 31.0%; Pred. No. 3.10e+00;
Matches 9; Conservative 12; Mismatches 7; Indels 1; Gaps 1;

Db 926 LSVFNPASNLVET-SLIISSTQASITSP 953

QY 62 LRVFDPSTGALVDSKSYAFSTNDTTSAA 90

RESULT 11
ID FOBI_YEAST STANDARD; PRT: 566 AA.
AC O13329; 004587;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE DNA REPLICATION FORK BLOCKING PROTEIN FOBI.
GN FOBI OR YDR110W OR YD9727.06.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97233188.
RA KOBAYASHI T., HORIUCHI T.;
RT "A yeast gene product, Fobi protein, required for both replication
RT fork blocking and recombinational hot spot activities.";
RT GENES CELLS 1:465-474(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA MURPHY L., SHORE L., HARRIS D., BARRELL B., RAJANDREAM M.A.,
RA WALSH S.V.;
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: ESSENTIAL FOR BOTH DNA REPLICATION FORK BLOCKING AND
CC RECOMBINATIONAL HOTSPOT ACTIVITIES.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF013245; G2331297; -
DR EMBL; Z48758; G747885; -
DR SGD; L0003959; FOBI.
FT CONFLICT 103 103 A -> T (IN REF. 2).
SQ SEQUENCE 566 AA; 65288 MW; 49D933EB CRC32;

Query Match 29.7%; Score 69; DB 1; Length 566;
Best Local Similarity 33.3%; Pred. No. 4.42e+00;
Matches 7; Conservative 8; Mismatches 5; Indels 1; Gaps 1;

Db 223 IHIEVFEPFNGEAEKGK-YSY 242

QY 60 LHLRVFDPSTGALVDSKSYAF 80

```

RESULT 12
ID PRVB_ESOLU STANDARD; PRT; 107 AA.
AC P02619;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PARVALBUMIN BETA.
OS ESOLX LUCIUS (NORTHERN PIKE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC TELEOSTEI; EUTELEOSTEI; PROTACANTHOPTERYGII; ESOCIFORMES; ESOCIDAE;
OC ESOLX.
RN [1]
RP SEQUENCE.
RX MEDLINE; 77090998.
RA GERDAY C.;
RT "The primary structure of the parvalbumin II of pike (Esox lucius).";
RL EUR. J. BIOCHEM. 70:305-318(1976).
RN [2]
RP CALCIUM-BINDING DATA.
RX MEDLINE; 84051312.
RA PERMYAKOV E.A., MEDVEDKIN V.N., KALINICHENKO L.P., BURSTEIN E.A.;
RT "Comparative study of physicochemical properties of two pike
parvalbumins by means of their intrinsic tyrosyl and phenylalanyl
fluorescence.";
RL ARCH. BIOCHEM. BIOPHYS. 227:9-20(1983).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.98 ANGSTROMS).
RX MEDLINE; 89011963.
RA DECLERQ J.-P., TINANT B., PARELLO J., ETIENNE G., HUBER R.;
RT "Crystal structure determination and refinement of pike 4.10
parvalbumin (minor component from Esox lucius).";
RL J. MOL. BIOL. 202:349-353(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).
RX MEDLINE; 91350177.
RA DECLERQ J.-P., RAMBAUD J.;
RT "Ionic interactions with parvalbumins. Crystal structure
determination of pike 4.10 parvalbumin in four different ionic
environments.";
RL J. MOL. BIOL. 220:1017-1039(1991).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (0.91 ANGSTROMS).
RA DECLERQ J.-P., EVRARD C., LAMZIN V., PARELLO J.;
RL SUBMITTED (OCT-1998) TO THE PDB DATA BANK.
CC -!- FUNCTION: IN MUSCLE, THE CALCIUM-BINDING PROTEIN PARVALBUMIN IS
THOUGHT TO BE INVOLVED IN MUSCLE RELAXATION.
CC -!- THIS IS ONE OF TWO MAJOR PARVALBUMINS FOUND IN THE WHITE MUSCLE
OF PIKE.
CC -!- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS, BELONGS TO
THE PARVALBUMINS SUBFAMILY.
DR PIR; A03054; PVPK2.
DR PDB; 1PAL; 15-JAN-92.
DR PDB; 2PAL; 15-JAN-92.
DR PDB; 3PAL; 15-JAN-92.
DR PDB; 4PAL; 15-JAN-92.
DR PDB; 2PVB; 07-OCT-98.
DR PROSITE; PS00018; EF_HAND; 2.
DR PFAM; PF00036; ehand; 2.
KW CALCIUM-BINDING; MUSCLE PROTEIN; REPEAT; ACETYLATION; 3D-STRUCTURE.
FT MOD_RES 1 1
FT CA_BIND 50 61
FT CA_BIND 89 100
FT TURN 2 4
FT HELIX 7 16
FT TURN 17 17
FT TURN 20 21
FT HELIX 25 32
FT TURN 33 33
FT TURN 35 36
FT TURN 39 49
FT TURN 51 52

```

```

FT STRAND 57 57
FT HELIX 59 63
FT TURN 64 64
FT HELIX 65 68
FT TURN 69 69
FT TURN 71 72
FT HELIX 78 88
FT TURN 90 91
FT STRAND 96 96
FT HELIX 98 105
FT TURN 106 106
SQ SEQUENCE 107 AA; 11390 MW; 1D1088D7 CRC32;

Query Match 29.3%; Score 68; DB 1; Length 107;
Best Local Similarity 32.4%; Pred. No. 6.29e+00;
Matches- 11; Conservative 9; Mismatches 14; Indels 0; Gaps 0;

Db 64 LFLQNEPSARALTDAETKAFADGDKDGMIG 97
| : | | | : | | : | | : | : | :
QY 60 LHLRVDPSTGALVDSKSYAFSTSDTTSAAFVS 93

RESULT 13
ID Y700_METJA STANDARD; PRT; 324 AA.
AC Q58111;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN MJ0700.
GN MJ0700.
OS METHANOCOCCUS JANNASCHII.
OC ARCHAEA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE;
OC METHANOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE; 96337999.
RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-P., ADAMS M.D., REICH C.I.,
RA OVERBEK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
RA SCOTT J.L., GEOHAGAN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii."
RL SCIENCE 273:1058-1073(1996).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: SOME, TO THE UPF0003 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67516; G1591415; -.
DR TIGR; MJ0700; -.
DR PFAM; PF00924; UPF0003; 1.
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 8 28
FT TRANSMEM 42 62
FT TRANSMEM 80 100
FT TRANSMEM 116 136
FT SEQUENCE 324 AA; 37260 MW; 7055857F CRC32;

Query Match 29.3%; Score 68; DB 1; Length 324;
Best Local Similarity 26.7%; Pred. No. 6.29e+00;
Matches 8; Conservative 14; Mismatches 6; Indels 2; Gaps 2;

```



```

ID OI1979 PRELIMINARY; PRT; 3443 AA.
AC OI1979;
DT 01-JUL-1997 (TREMBLEL. 04, CREATED)
DT 01-JUL-1997 (TREMBLEL. 04, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLEL. 08, LAST ANNOTATION UPDATE)
DE POLYPROTEIN.
OS MAIZE CHLOROTIC DWARF WAIKAVIRUS.
OC VIRUSES: SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; SEQUIVIVIRIDAE;
OC WAIKAVIRUS.
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=TENNESSEE (TN);
RX MEDLINE; 97296968.
RA REDDICK B.B., HABERA L.F., LAW M.D.;
RT "Nucleotide sequence and taxonomy of maize chlorotic dwarf virus within the family Sequiviridae.";
RL J. GEN. VIROL. 78:1165-1174(1997).
DR ENBL; U67839; G2160664; -.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
KW POLYPROTEIN.
SQ SEQUENCE 3443 AA; 387659 MW; 7560C323 CRC32;

Query Match 34.5%; Score 80; DB 14; Length 3443;
Best Local Similarity 32.4%; Pred. No. 2.01e-01;
Matches 11; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

Db 1201 GGVLRIWHDPGSLDEGVFAMSTNLPPPGAFV 1234
QY |:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~::~|||
59 GLHLRVDFPSTGALVDKSFAFTSNDTTSAAVF 92

RESULT 3
ID Q47628 PRELIMINARY; PRT; 424 AA.
AC Q47628;
DT 01-NOV-1996 (TREMBLEL. 01, CREATED)
DT 01-NOV-1996 (TREMBLEL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLEL. 08, LAST ANNOTATION UPDATE)
DE TIEB PROTEIN.
GN SENB.
OS ESCHERICHIA COLI.
OC BACTERIA: PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERiaceae;
OC ESCHERICHIA.
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=ENTEROINVASIVE;
RX MEDLINE; 96071891.
RA NATARO J., SERIWATANA J., FASANO A., MANEVAL D., GUERS L., NORIEGA M.,
RA DUBOVSKY F., LEVINE M.M., MORRIS J.G.;
RT "Identification and cloning of a novel plasmid-encoded enterotoxin of
entero-invasive Escherichia coli and Shigella strains.";
RL INFECT. IMMUN. 63:4721-4728(1995).
DR ENBL; Z54195; E198841; -.
SQ SEQUENCE 424 AA; 47958 MW; 2366BD23 CRC32;

Query Match 32.8%; Score 76; DB 2; Length 424;
Best Local Similarity 39.4%; Pred. No. 8.68e-01;
Matches 13; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

Db 151 SASLRISMFDPTPKGLVRSDRLSFAGNAADS 183
QY ||::|:|:|:|:|:|:|:|:|:|:|:|:|:|:~::~|||
57 SAGHLRVDFPSTGALV-DKSYAFTSNDDTs 88

RESULT 4
ID Q68542 PRELIMINARY; PRT; 1449 AA.
AC Q68542;
DT 01-AUG-1998 (TREMBLEL. 07, CREATED)
DT 01-AUG-1998 (TREMBLEL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLEL. 07, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE N (FRAGMENT).
GN GTFN.
OS STREPTOCOCCUS SALIVARIUS.
```

```

RX MEDLINE; 95095028.
RA RUDANT E., BOURLIOUX P., COURVALIN P., LAMBERT T.;
RT "Characterisation of the aac(6')-Ik gene of Acinetobacter sp. 6.";
RL FEMS MICROBIOL. LETT. 124:49-54(1994).
DR EMBL; L29510: G847647; -.
DR PFAM; PF00583; Acetyltransf; 1.
SQ SEQUENCE 145 AA; 16638 MW; C824BA11 CRC32;

Query Match      31.0%; Score 72; DB 2; Length 145;
Best Local Similarity 30.6%; Pred. No. 3.58e+00;
Matches 11; Conservative 10; Mismatches 13; Indels 2; Gaps 1;

Db 42 ALQLLVSDHVGMLSEASIRYVNGTETSPVAFL 77
   :| | : | | | | | | | | | | | | | |
Qy 59 GLHLRVF--DPSTGALVDKSKSYAFSTNDTTSAVF 92

RESULT          9
ID Q13641        PRELIMINARY;           PRT;    420 AA.
AC Q13641;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE 574 ONCOFETAL ANTIGEN PRECURSOR.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HUMO.
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=PLACENTA;
RC MEDLINE; 94179356.
RX MYERS K.A., RAHI-SAUND V., DAVISON M.D., YOUNG J.A., CHEATER A.J.,
RA STERN P.L.;
RT "Isolation of a cDNA encoding 574 oncofetal trophoblast glycoprotein.
RT An antigen associated with metastasis contains leucine-rich
RT repeats."
RL J. BIOL. CHEM. 269:9319-9324(1994).
DR EMBL; Z29083; G435655; -.
DR PFAM; PF00560; LRR; 4.
KW SIGNAL.
FT SIGNAL          1 31 POTENTIAL
FT CHAIN          32 420 574 ONCOFETAL ANTIGEN.
SQ SEQUENCE 420 AA; 46031 MW; 43633112 CRC32;

Query Match      31.0%; Score 72; DB 4; Length 420;
Best Local Similarity 44.8%; Pred. No. 3.58e+00;
Matches 13; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

Db 144 LRQLDLSHNPLADLSPFPFGSGSNASVAP 172
   || | | : | | : || | | | | : | |
Qy 62 LRVDFPSTGALVDKSKSYAFSTNDTTSA 90

RESULT         10
ID Q22056        PRELIMINARY;           PRT;    572 AA.
AC Q22056;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PLASTID RNA POLYMERASE SIGMA-SUBUNIT.
GN SIG1 OR SIGA OR SIGB.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPLANTEE; STREPTOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=COLUMBIA; TISSUE=LEAF;
RC MEDLINE; 98070858.
RX ISONO K., MASANORI S., YOSHIMOTO K., NIWA Y., SATOH K., YOKOTA A.,
RA KOBAYASHI H.;
RT "Leaf-specifically expressed genes for polypeptides destined for
RT chloroplasts with domains of sigma70 factors of bacterial RNA
```

RT polymerases in Arabidopsis thaliana.";
RL PROC. NATL. ACAD. SCI. U.S.A. 94:14948-14953(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA HAKIMI A., PRIVAT I., LERBS-MACHE S.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-COLUMBIA;
RA TANAKA K., TOZAWA Y., MOCHIZUKI N., SHINOZAKI K., NAGATANI A.,
RA WAKASA K., TAKAHASHI H.;
RT "Characterization of three cDNA species encoding plastid RNA
polymerase sigma factors in Arabidopsis thaliana: evidence for the
sigma factor heterogeneity in higher plant plastids.";
RL FEBS LETT. 413:309-313(1997).
RN [4]
RP SEQUENCE FROM N.A.
RA SHINN P., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y., SUN H.,
RA CONWAY A., CONWAY A., KURTZ D., OUI O., SHEN Y.K., TORIUMI M.,
RA VYOTSKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THEOLOGIS A.,
RA ECKER J.R.;
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AB004820; D1025747; -;
DR EMBL; Y13622; E1169752; -;
DR EMBL; AB004293; D1023292; -;
DR EMBL; AC003981; G3063440; -;
DR PFAM; PF00140; sigma70; 1.
KW TRANSIT PEPTIDE; SIGMA FACTOR.
SQ SEQUENCE 572 AA; 64054 MW; C0B0407E CRC32;

Query Match 31.0%; Score 72; DB 10; Length 572;
Best Local Similarity 19.4%; Pred. No. 3.58e+00;
Matches 6; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

Db 58 LRLPFDSDSLSDROWTYTRPDGPSTEA 88
QY 60 LHLRFDPSTGALVDSKSYAFSTSNDDTSA 90

RESULT 11
ID 022455 PRELIMINARY; PRT; 574 AA.
AC 022455;
DT 01-JAN-1998 (TREMREL. 05, CREATED)
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE SIGMA FACTOR 2.
GN SIG.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA, VIRIDIPANTAE; STREPTOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-COLUMBIA; TISSUE=3 DAY-OLD SEEDLING HYPOCOTYL;
RA ALLISON L.A.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF015543; G2353173; -;
DR PFAM; PF00140; sigma70; 1.
SQ SEQUENCE 574 AA; 64323 MW; 40F28766 CRC32;

Query Match 31.0%; Score 72; DB 10; Length 574;
Best Local Similarity 19.4%; Pred. No. 3.58e+00;
Matches 6; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

Db 58 LRLPFDSDSLSDROWTYTRPDGPSTEA 88
QY 60 LHLRFDPSTGALVDSKSYAFSTSNDDTSA 90

RESULT 12
ID 012071 PRELIMINARY; PRT; 889 AA.
AC 012071;

DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMREL. 01, LAST ANNOTATION UPDATE)
DE ORF:PF889.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RA EIDE L.G., SANDER C., PRYDZ H.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA BOWMAN S.;
RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA BARRELL B., RAJANDREAM M.A.;
RL SUBMITTED (JAN-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RA PRYDZ H., EIDE L.G.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [5]
RP SEQUENCE FROM N.A.
RA MIPS;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X95966; E225338; -;
DR EMBL; Z47814; G642299; -;
DR EMBL; Z74323; E253298; -;
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 889 AA; 101519 MW; 3C15605E CRC32;

Query Match 31.0%; Score 72; DB 3; Length 889;
Best Local Similarity 31.0%; Pred. No. 3.58e+00;
Matches 9; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

Db 77 RSEDSRTLSVANSRSGFENETHSGSND 105
QY 63 RVDFPSTGALVDSKSYAFSTSNDDTSA 91

RESULT 13
ID 026845 PRELIMINARY; PRT; 241 AA.
AC 026845;
DT 01-JAN-1998 (TREMREL. 05, CREATED)
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE CONSERVED PROTEIN.
GN MTH751.
OS METHANOBACTERIUM THERMOAUTOTROPHICUM.
OC ARCHAEA; EURYARCHAEOTA; METHANOBACTERIALES; METHANOBACTERIACEAE;
OC METHANOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H;
RX MEDLINE; 98037514.
RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,
RA ALDRIDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,
RA HARRISON D., HOANG L., KEAGLE P., LUMM W., POTIER B., QIU D.,
RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,
RA JIWANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S.,
RA MCDUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,
RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;
RT *Complete genome sequence of Methanobacterium thermoautotrophicum
J. BACTERIOL. 179:7135-7155(1997).
RL EMBL; AE000854; G2621840; -;
SQ SEQUENCE 241 AA; 26734 MW; E6584334 CRC32;

Query Match 30.6%; Score 71; DB 1; Length 241;

Best Local Similarity 38.9%; Pred. No. 5.07e+00;
Matches 14; Conservative 8; Mismatches 11

Matches 14; Conservative 8; Mismatches 11; Indels 3; Gaps 2;

```

Db      24  AGMNIHI-DPGPGALV--RSYQFDSDPRLDAVMVS 56
      ||::: ||: |||| :|| | : | ||
Qy      58  AGLHLRVFDPSTGALVDSKSYAFSTSDNTTSAAFVS 93

```

[illegible]

Query Match 30.6%; Score 71; DB 14; Length 351;
Best Local Similarity 34.5%; Pred. No. 5.07e+00;
Matches 10; Conservative 10; Mismatches 7; Indels

Matches	10;	Conservative	10;	Mismatches	7;	Indels	2;	Gaps	2;
----------------	-----	---------------------	-----	-------------------	----	---------------	----	-------------	----

```

Db      259 FE-ATGNLVAPR-YAFAMNRGSGGIIIS 285
      | : | | | : | | | : | : |
Qy      65 FDPSTGALVDSKSYAFSTNDTSAAEVS 93

```

RESULT	15	
ID	Q67091	PRELIMINARY; PRT; 357 AA.
AC	Q67091;	
DT	01-NOV-1996	(TREMBREL. 01, CREATED)
DT	01-NOV-1996	(TREMBREL. 01, LAST SEQUENCE UPDATE)
DT	01-NOV-1998	(TREMBREL. 08, LAST ANNOTATION UPDATE)
DE		Hemagglutinin precursor.
DN	HA.	
OS	INFLUENZA A VIRUS.	
OC	VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; ORTHOMYXOVIRIDAE;	
ON	INFLUENZA VIRUS A AND B GROUP.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-A/SWINE/HK/WEHIVE/1/80;	
RA	SUGITA S., YOSHIOKA Y., ITAMURA S., KANEGAE Y., OGUCHI K.	
RA	GOJOBORI T., NEROME K., OYA A. ;	
RJ	J. MOL. EVOL. 32:16-23(1994).	
DR	EMLB; X57494; G516377; "	
DR	PFAM; PF00509; Hemagglutinin; 1.	
KW	SIGNAL; HEMAGGLUTININ	
FT	SIGNAL	1
FT	CHAIN	17
FT	CHAIN	18 343
FT	CHAIN	345 357
FT	CHAIN	357 AA; 39544 MW; 9C2D9A2F CRC32.
SO	SEQUENCE	POTENTIAL. HEMAGGLUTININ. HEMAGGLUTININ.

Query Match 30.6%; Score 71; DB 14; Length 357;
Best Local Similarity 37.9%; Pred. No. 5.07e+00;
Matches 11; Conservative 9; Mismatches 7; Indels

Matches	11; Conservative	9; Mismatches	7; Indels	2; Gaps	2;
---------	------------------	---------------	-----------	---------	----

```

Db      259 FE-ATGNLVVPR-YAFAMNRDPGSGIIIS 285
      | : || | | : ||| : | | : ||
Qv      65 FDPSTGALVDKSYAFSTSNDDTTSAAFVS 93

```

Search completed: Wed Sep 1 16:24:42 1999

M P S R C H
***** (TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Sep 1 16:28:14 1999; MasPar time 3.49 Seconds
Tabular output not generated. 48.772 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (57-64) from PCTUS9913024.pep (10 of 12)
Perfect Score: 56
Sequence: 1 SAGLHLRV 8

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 15.625; Variance 42.633; scale 0.366

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	47	83.9	28	7	R34572	Domain 1 from integra
2	45	80.4	185	32	W60046	Human TNF receptor re
3	45	80.4	240	21	W05810	Human tumour necrosis
4	44	78.6	467	36	W76426	Glucocorticoid induci
5	44	78.6	467	36	W76427	Rat Glucocorticoid in
6	44	78.6	843	3	R14924	Mutant SP6DNA polymer
7	44	78.6	844	3	R14927	Mutant SP6DNA polymer
8	44	78.6	849	3	R14922	SP6DNA polymerase
9	44	78.6	849	3	R14923	Mutant SP6DNA polymer
10	44	78.6	849	3	R14925	Mutant SP6DNA polymer
11	44	78.6	849	3	R14926	Mutant SP6DNA polymer
12	43	76.8	523	1	P80674	NtrA gene product.
13	43	76.8	1891	26	W22610	Platenolide synthase
14	43	76.8	1891	26	W23720	Platenolide synthase
15	42	75.0	112	34	W77578	Staphylococcus aureus
16	42	75.0	172	16	R82811	Rhnpicephalus appendi

17	42	75.0	438	25	W18209	Staphylococcus aureus	2.66e+02
18	42	75.0	655	6	R31041	smrR polypeptide.	2.66e+02
19	42	75.0	1073	28	W32063	Human ST receptor pro	2.66e+02
20	42	75.0	1073	28	W37371	Human ST receptor pro	2.66e+02
21	42	75.0	1075	8	R38861	GC-C.	2.66e+02
22	42	75.0	1349	33	W59359	Human retinal degene	2.66e+02
23	42	75.0	1422	16	R82066	Hepatitis GB virus (H	2.66e+02
24	42	75.0	1706	16	R92103	Rat RIZ.	2.66e+02
25	41	73.2	30	37	W80945	Amino acid sequence o	3.45e+02
26	41	73.2	33	37	W80946	Amino acid sequence o	3.45e+02
27	41	73.2	108	18	W00649	G-CSF receptor ligand	3.45e+02
28	41	73.2	213	18	W00647	Mouse G-CSF receptor	3.45e+02
29	41	73.2	255	18	R99142	Ligand binding cytoke	3.45e+02
30	41	73.2	299	39	W80984	Amino acid sequence o	3.45e+02
31	41	73.2	299	39	W80985	B cell receptor assoc	3.45e+02
32	41	73.2	299	34	W61629	Clone HMWS46 of Proh	3.45e+02
33	41	73.2	334	18	R99140	Ligand binding cytoke	3.45e+02
34	41	73.2	412	37	W80942	Human heart lecithin-	3.45e+02
35	41	73.2	444	37	W80943	Human kidney lecithin-	3.45e+02
36	41	73.2	480	17	R85310	Cystathionine gamma s	3.45e+02
37	41	73.2	509	17	R85311	Cystathionine gamma s	3.45e+02
38	41	73.2	837	3	R14254	Murine GCSF receptor	3.45e+02
39	40	71.4	308	20	W10666	Human TGF-beta-like c	4.47e+02
40	40	71.4	308	20	W10670	Human TGF-beta-like c	4.47e+02
41	40	71.4	308	20	W10673	Human TGF-beta-like c	4.47e+02
42	40	71.4	308	16	R84710	Human embryonal lung	4.47e+02
43	40	71.4	323	1	R05711	Tumour-associated ant	4.47e+02
44	40	71.4	4303	17	R90302	Polycystic kidney dis	4.47e+02
45	40	71.4	4339	15	R75916	Polycystic kidney dis	4.47e+02

ALIGNMENTS

RESULT 1
ID R34572 standard; Protein; 28 AA.
AC R34572;
DE 04-JUN-1993 (first entry)
DE Domain 1 from integrase like protein Tn2603 ORF3.
KW cytotoxic T-lymphocyte response; transformed Mycobacteria; BCG;
KW Mycobacterium smegmatis; vaccine; cell mediated immunity; HIV;
PN pertussis; malaria; influenza virus; CTL; herpes virus.
PD W09307897-A.
PF 29-APR-1993.
PR 21-OCT-1992; U09075.
PR 21-OCT-1991; US-780261.
PA (MEDI-) MEDIMUNE INC.
PI Stover CK;
DR WPI; 93-152187/18.
PT Expression vector for expressing protein or polypeptide in
PT mycobacterium - contg DNA sequences encoding lipoprotein
PT secretion signal and peptide heterologous to bacteria expressing
PT fusion protein of lipoprotein heterologous to bacteria
PS Disclosure; Fig 29; 86pp; English.
CC This sequence is domain 1 from the integrase like protein Tn2603 ORF 3.
CC The whole coding sequence of the enzyme could be included in an
CC expression vector in order to aid its integration into its host. The
CC expression vector also includes at least the secretion signal of a
CC lipoprotein and a second sequence encoding a heterologous protein
CC and a mycobacterial promoter to control expression of the
CC heterologous protein.
SQ Sequence 28 AA;
Query Match 83.9%; Score 47; DB 7; Length 28;
Best Local Similarity 75.0%; Pred. No. 6.96e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 15 segqlgrv 22
| | | | |
QY 57 SAGLHLRV 64

RESULT 2
ID W60046 standard; Protein; 185 AA.

AC W60046;
DT 25-SEP-1998 (first entry)
DE Human TNF receptor related splice variant 1 (TR2-SV1) protein.
KW Human: tumour necrosis factor; TNF; herpes simplex virus;
KW aberrant cell survival; radiation therapy; lymphocyte proliferation;
KW immune deficiency syndrome.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..36
FT /note= "Signal peptide"
FT 37..185
FT /note= "Mature peptide"
FT W09818824-A1.
PN 07-MAY-1998.
PD 30-OCT-1996; U18540.
PR 30-OCT-1996; WO-U18540.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Gentz RL, Hurle MR, Lyn SDP, Ni J, Rosen CA;
DR WPI: 98-272139/24.
DR N-PSDB; V34510.
DR Nucleic acid encoding TR2 tumour necrosis factor family receptor -
PT and its splice variants, useful for diagnosis and treatment of
PT diseases involving abnormal cell survival or death, e.g. herpes
PT simplex infection
PS Claim 1: Fig 4; 151pp; English.
CC The human tumour necrosis factor (TNF) receptor related protein (TR2) is
CC a member of the TNF family and displays considerable homology to murine
CC CD40. It can be used in soluble forms to treat herpes simplex virus
CC infection and TR2 proteins (or their agonists or antagonists) are used to
CC treat disease associated with aberrant cell survival. Agonists may also
CC be used to protect against the effects of radiation therapy and to
CC stimulate lymphocyte proliferation and differentiation in patients
CC infected by human immune deficiency syndrome.
SQ Sequence 185 AA;

Query Match 80.4%; Score 45; DB 32; Length 185;
Best Local Similarity 100.0%; Pred. No. 1.20e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 169 aglhlr 174
QY 58 AGLHLR 63
|||||
RESULT 3
ID W05810 standard; Protein; 240 AA.
AC W05810;
DT 30-JUN-1997 (first entry)
DE Human tumour necrosis factor receptor splice variant 1.
KW Human: tumour necrosis factor; TNF; receptor; treatment;
KW activation; inhibition; identification; agonist; antagonist;
KW stimulation; T cell; differentiation; mediation; immune;
KW antiviral; response; regulation; growth; protection; HIV;
KW radiation; Chlamydia; infection; immunodeficiency; virus;
KW autoimmune disease; inflammation; septic shock; cerebral; malaria;
KW cachexia; B cell; cancer; graft; host; reaction; rejection;
KW prevention; apoptosis; cytotoxicity; hybridisation probe;
KW detection; antibody; reagent; splice variant 1; SV1.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..38
FT /label= sig_peptide
FT 39..240
FT /label= mat_peptide
FT W09634095-A1.
PN 31-OCT-1996.
PD 27-APR-1995; U05058.
PR 27-APR-1995; WO-U05058.
PA (HUMA-) HUMAN GENOME SCI INC.
PA Gentz R, Ni J, Rosen CA;
DR WPI: 96-497627/49.
DR N-PSDB; T40049.

PT New nucleic acid encoding a human tumour necrosis factor receptor -
PT useful for treatment of auto-immune diseases etc., in diagnosis and
PT for drug screening
PS Claim 1: Pages 48-49: 73pp; English.
CC The present sequence is the human tumour necrosis factor (TNF)
CC receptor splice variant 1 (SV1), which may be used for the
CC identification of TNF SV1 receptor agonists or antagonists. TNF SV1
CC receptor agonists inhibit tumour growth, stimulate cell (e.g.
CC T cell) differentiation, mediate the immune and antiviral
CC responses, regulate growth, protect against radiation and Chlamydia
CC infection, and can be used to treat immunodeficiencies, e.g. human
CC immunodeficiency virus. Antagonists can be used to treat T cell
CC mediated autoimmune disease, inflammation, septic shock, cerebral
CC malaria, cachexia or B cell cancers, to inhibit graft-host
CC reactions and to prevent apoptosis or cytotoxicity. Fragments of
CC the TNF SV1 receptor encoding DNA can be used as hybridisation
CC probes for detecting related genes. Antibodies against the protein
CC can be used as reagents for detecting/measuring soluble forms of
CC protein in the circulation.
SQ Sequence 240 AA;

Query Match 80.4%; Score 45; DB 21; Length 240;
Best Local Similarity 100.0%; Pred. No. 1.20e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 169 aglhlr 174
QY 58 AGLHLR 63
|||||
RESULT 4
ID W76426 standard; Protein; 467 AA.
AC W76426;
DT 15-JAN-1999 (first entry)
DE Glucocorticoid inducible suppressor protein.
KW Glucocorticoid inducible suppressor protein; interleukin-8; rheumatism;
KW intracellular signal transduction inhibitor; IL-8 promoter; allergy;
KW inflammatory disease; bronchial asthma; therapy.
OS Homo sapiens.
PN W09838213-A1.
PD 03-SEP-1998.
PF 27-FEB-1998; J00836.
PR 28-FEB-1997; JP-062008.
PR (CYTO-) INST CYTOSIGNAL RES INC.
PI Ohtsuka T, Yoshida H;
DR WPI: 98-481142/41.
DR N-PSDB; V61810.
PT Protein inhibiting activation of interleukin-8 promoter in response
PT to extracellular stimulus - for use in diseases involving
PT Interleukin-8 expression, such as inflammatory disorders and asthma
PS Claim 1: Page 39-43; 72pp; Japanese.
CC This sequence represents the human glucocorticoid inducible suppressor
CC protein (GISP) of the invention. The GISP is an intracellular signal
CC transduction inhibitor, and inhibits the activation of interleukin-8
CC (IL-8) promoter activity in response to a specific extracellular stimulus
CC (especially by interleukin-1 beta). The invention can be used in the
CC treatment and prevention of disorders in which IL-8 expression is
CC involved, such as inflammatory diseases, bronchial asthma, allergy and
CC rheumatism.
SQ Sequence 467 AA;

Query Match 78.6%; Score 44; DB 36; Length 467;
Best Local Similarity 71.4%; Pred. No. 1.56e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 32 aglhlr 38
QY 57 SAGLHLR 63
::|||
RESULT 5
ID W76427 standard; Protein; 467 AA.
AC W76427;

DT 15-JAN-1999 (first entry)
 DE Rat Glucocorticoid inducible suppressor protein.
 KW Glucocorticoid inducible suppressor protein; Interleukin-8; rheumatism;
 KW Intracellular signal transmission inhibitor; IL-8 promoter; allergy;
 KW inflammatory disease; bronchial asthma; therapy.
 OS Rattus sp.
 PN WO9838213-A1.
 PD 03-SEP-1998.
 PF 27-FEB-1998; J00836.
 PR 28-FEB-1997; JP-062008.
 PA (CYTO-) INST CYTOSIGNAL RES INC.
 PI Ohtsuka T, Yoshida H.
 DR WPI; 98-481142/41.
 DR N-PSDB; V61811.
 PT Protein inhibiting activation of interleukin-8 promoter in response
 PT to extracellular stimulus - for use in diseases involving
 PT interleukin-8 expression, such as inflammatory disorders and asthma
 PS Claim 3; Page 46-50; 72pp; Japanese.
 CC This sequence represents the rat glucocorticoid inducible suppressor
 CC protein (GISP) of the invention. The GISP is a intracellular signal
 CC transmission inhibitor, and inhibits the activation of interleukin-8
 CC (IL-8) promoter activity in response to a specific extracellular stimulus
 CC (especially by interleukin-1 beta). The invention can be used in the
 CC treatment and prevention of disorders in which IL-8 expression is
 CC involved, such as inflammatory diseases, bronchial asthma, allergy and
 CC rheumatism.
 SQ Sequence 467 AA;

Query Match 78.6%; Score 44; DB 3; Length 467;
 Best Local Similarity 71.4%; Pred. No. 1.56e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 31 apghlhr 37
 :|||||
 QY 57 SAGLHLRV 63

RESULT 6

ID R14924 standard; Protein; 843 AA.
 AC R14924;
 DT 11-FEB-1992 (first entry)
 DE Mutant SP6DNA polymerase.
 KW Exonuclease.
 OS SP6 phage.
 PN J03232487-A.
 PD 16-OCT-1991.
 PF 06-FEB-1990; 025185.
 PR 06-FEB-1990; JP-025185.
 PA (TAKA-) TAKARA SHUZO KK.
 DR WPI; 91-349003/48.
 PT SP6DNA polymerase of defined sequence - and modified polymerase
 PT lacking 3',5'-exo-nuclease activity
 PS Disclosure; Page 8; 13pp; Japanese
 CC The amino acid sequence is that of mutated SP6DNA polymerase, where
 CC amino acids His167-His172 have been deleted by site specific
 CC mutation. This results in a mutant lacking 3',5'-exonuclease
 CC activity. See also R14922-R14927.
 SQ Sequence 843 AA;

Query Match 78.6%; Score 44; DB 3; Length 843;
 Best Local Similarity 75.0%; Pred. No. 1.56e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 565 gaglelrv 572
 :|||||
 QY 57 SAGLHLRV 64

RESULT 7

ID R14927 standard; Protein; 844 AA.
 AC R14927;
 DT 11-FEB-1992 (first entry)
 DE Mutant SP6DNA polymerase.

KW Exonuclease.
 OS SP6 phage.
 PN J03232487-A.
 PD 16-OCT-1991.
 PF 06-FEB-1990; 025185.
 PR 06-FEB-1990; JP-025185.
 PA (TAKA-) TAKARA SHUZO KK.
 DR WPI; 91-349003/48.
 PT SP6DNA polymerase of defined sequence - and modified polymerase
 PT lacking 3',5'-exo-nuclease activity
 PS Disclosure; Page 8; 13pp; Japanese
 CC The amino acid sequence is that of mutated SP6DNA polymerase, where
 CC amino acids Val164-Ser168 have been deleted by site specific
 CC mutation. See also R14922-R14926.
 SQ Sequence 844 AA;

Query Match 78.6%; Score 44; DB 3; Length 844;
 Best Local Similarity 75.0%; Pred. No. 1.56e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 566 gaglelrv 573
 :|||||
 QY 57 SAGLHLRV 64

RESULT 8

ID R14922 standard; Protein; 849 AA.
 AC R14922;
 DT 11-FEB-1992 (first entry)
 DE SP6DNA polymerase.
 KW Exonuclease.
 OS SP6 phage.
 PN J03232487-A.
 PD 16-OCT-1991.
 PF 06-FEB-1990; 025185.
 PR 06-FEB-1990; JP-025185.
 PA (TAKA-) TAKARA SHUZO KK.
 DR WPI; 91-349003/48.
 DR N-PSDB; Q14772.
 PT SP6DNA polymerase of defined sequence - and modified polymerase
 PT lacking 3',5'-exo-nuclease activity
 PS Claim 1; Fig 1; 13pp; Japanese
 CC The amino acid sequence is that of SP6DNA polymerase and was obtd.
 CC from the SP6 phage. It can be produced recombinantly and is of
 CC property, superior to conventional polymerases. See also R14923-
 CC R14927
 SQ Sequence 849 AA;

Query Match 78.6%; Score 44; DB 3; Length 849;
 Best Local Similarity 75.0%; Pred. No. 1.56e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 571 gaglelrv 578
 :|||||
 QY 57 SAGLHLRV 64

RESULT 9

ID R14923 standard; Protein; 849 AA.
 AC R14923;
 DT 11-FEB-1992 (first entry)
 DE Mutant SP6DNA polymerase.
 KW Exonuclease.
 OS SP6 phage.
 PN J03232487-A.
 PD 16-OCT-1991.
 PR 06-FEB-1990; 025185.
 PA (TAKA-) TAKARA SHUZO KK.
 DR WPI; 91-349003/48.
 PT SP6DNA polymerase of defined sequence - and modified polymerase
 PT lacking 3',5'-exo-nuclease activity
 PS Disclosure; Page 8; 13pp; Japanese

CC The amino acid sequence is that of mutated SP6DNA polymerase, where
 CC the Asp at position 101 has been changed by site specific mutation
 CC to Ala. See also R14922-R14927.

SQ Sequence 849 AA;

Query Match 78.6%; Score 44; DB 3; Length 849;

Best Local Similarity 75.0%; Pred. No. 1.56e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 571 gaglelrv 578

QY 57 SAGLHLRV 64

RESULT 10

ID R14925 standard; Protein; 849 AA.

AC R14925;

DT 11-FEB-1992 (first entry)

DE Mutant SP6DNA polymerase.

KW Exonuclease.

OS SP6 phage.

PN J03232487-A.

PD 16-OCT-1991.

PF 06-FEB-1990; 025185.

PR 06-FEB-1990; JP-025185.

PA (TAKA-) TAKARA SHUZO KK.

DR WPI; 91-349003/48.

PT SP6DNA polymerase of defined sequence - and modified polymerase

PT lacking 3', 5'-exo-nuclease activity

PS Disclosure; Page 8; 13pp; Japanese

CC The amino acid sequence is that of mutated SP6DNA polymerase, where

CC the His at position 167 has been changed by site specific mutation

CC to Glu, Glu170 to Lys and His172 to Asp. See also R14922-R14927.

SQ Sequence 849 AA;

Query Match 78.6%; Score 44; DB 3; Length 849;

Best Local Similarity 75.0%; Pred. No. 1.56e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 571 gaglelrv 578

QY 57 SAGLHLRV 64

RESULT 11

ID R14926 standard; Protein; 849 AA.

AC R14926;

DT 11-FEB-1992 (first entry)

DE Mutant SP6DNA polymerase.

KW Exonuclease.

OS SP6 phage.

PN J03232487-A.

PD 16-OCT-1991.

PF 06-FEB-1990; 025185.

PR 06-FEB-1990; JP-025185.

PA (TAKA-) TAKARA SHUZO KK.

DR WPI; 91-349003/48.

PT SP6DNA polymerase of defined sequence - and modified polymerase

PT lacking 3', 5'-exo-nuclease activity

PS Disclosure; Page 8; 13pp; Japanese

CC The amino acid sequence is that of mutated SP6DNA polymerase, where

CC the His at position 167 has been changed by site specific mutation

CC to Gln and Ser168 to Asn. See also R14922-R14927.

SQ Sequence 849 AA;

Query Match 78.6%; Score 44; DB 3; Length 849;

Best Local Similarity 75.0%; Pred. No. 1.56e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 571 gaglelrv 578

QY 57 SAGLHLRV 64

RESULT 12

ID P80674 standard; Protein; 523 AA.

AC P80674;

DT 24-OCT-1990 (first entry)

DE NtrA gene product.

KW NtrA; sigma factor; transcriptional activation; nitrogen assimilation;

KW fixation; nif operon; C4-dicarboxylate transport.

OS Rhizobium meliloti.

PN EP-292984-A.

PD 30-NOV-1988.

PF 27-MAY-1988; 108482.

PR 29-MAY-1987; US-055228.

PA (GEHO-) Gen Hospital Corp.

PI Ronson C, Ausubel F;

DR WPI; 88-339561/48.

DR N-PSDB; N81266.

PT Cloned rhizobium meliloti ntrA gene - for use in C4-dicarboxylate

PT transport, nitrate assimilation, symbiotic nitrogen fixation and

PS Identifying pathogenic genes.

CC Claim 1; Page 12; 19pp; English.

CC The ntrA gene product is a transcriptional activator. It controls the

CC processes of nitrate assimilation, symbiotic nitrogen fixation and

CC C4-dicarboxylate transport, working in conjunction with NtrC, NifA and

CC DctD resp. It can also be used to identify genes involved in

CC pathogenesis.

CC See also P82126.

SQ Sequence 523 AA;

Query Match 76.8%; Score 43; DB 1; Length 523;

Best Local Similarity 85.7%; Pred. No. 2.04e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 4 saslhrlr 10

QY 57 SAGLHLR 63

RESULT 13

ID W22610 standard; Protein; 1891 AA.

AC W22610;

DT 27-FEB-1998 (first entry)

DE Platenolide synthase ORF5 protein.

KW Tylactone synthase gene cluster; tylG gene; multifunctional protein;

KW polyketide; tyllactone synthetis; antibiotic; tylosin.

OS Streptomyces ambofaciens.

PH Key Location/Qualifiers

FT Domain 35..458

FT /note= "ketosynthase domain, KS7"

FT Domain 582..917

FT /note= "acyltransferase domain, AT7"

FT Domain 1233..1418

FT /note= "ketoreductase domain, KR7"

FT Domain 1502..1585

FT /note= "acyl carrier protein domain, ACP7"

FT Domain 1715..1747

FT /note= "thioesterase domain, TE7"

PN EP-791655-A2.

PD 27-AUG-1997.

PF 19-FEB-1997; 301056.

PR 22-FEB-1996; US-012078.

PA (EJIL) LILLY & CO ELI.

PI Dehoff BS, Kuhstoss SA, Rostock PR, Sutton KL;

DR WPI; 97-418046/39.

DR N-PSDB; T80414.

PT DNA encoding Streptomyces fradiae tyllactone synthase domain - for

PT production of tylosin-related polyketide compounds

PS Example 2; Pages 172-178; 220pp; English.

CC W22606-W22610 represent proteins encoded by the platenolide synthase gene

CC cluster. The gene cluster is also referred to as the srmg gene, and was

CC isolated from Streptomyces ambofaciens. These sequences are

CC multi-functional proteins which direct the synthesis of the polyketide

CC platenolide. Platenolide is the basic building block of the macrolide

CC antibiotic spiramycin. The DNA encoding this sequence was used along with
 CC the tylg gene (see T80413) to create a hybrid ORF1 sequence (see T80415).
 CC The tylg gene is the tyllactone synthase gene cluster of the invention.
 CC The tylg sequence was isolated from Streptomyces fradiae, and encodes
 CC multifunctional proteins which direct the synthesis of the polyketide
 CC tyllactone. Tyllactone is the basic building block of the antibiotic
 CC tyllactone. The hybrid sequence can be used to transform S. ambofaciens
 CC lacking the srmg ORF1 sequence, or S. fradiae lacking the tylg ORF1
 CC sequence, so that they can produce polyketides. The DNA sequence can be
 CC modified so as to alter the type of carboxylic acids incorporated, the
 CC number of carboxylic acids incorporated and/or the post-condensation
 CC reactions performed, thereby resulting in novel tyllactone-related
 CC polyketides.
 CC Sequence 1891 AA;

Query Match 76.8%; Score 43; DB 26; Length 1891;
 Best Local Similarity 75.0%; Pred. No. 2.04e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 1007 taglhryv 1014
 :|||||
 QY 57 SAGLHLRV 64

RESULT 14
 ID W23720 standard; Protein; 1891 AA.
 AC W23720;
 DT 27-FEB-1998 (first entry)
 DE Platenolide synthase ORF5 protein.
 KW Platenolide synthase gene cluster; platenolide production; srmg gene;
 KW multi-functional protein; macrolide antibiotic; spiramycin.
 OS Streptomyces ambofaciens.
 FH Key Location/Qualifiers
 FT Domain 35..458
 FT Domain 582..917
 FT Domain 1235..1418
 FT Domain /note= "acyltransferase domain, A77"
 FT Domain /note= "ketoreductase domain, KR7"
 FT Domain 1502..1585
 FT Domain /note= "acyl carrier protein domain, ACP7"
 FT Domain 1715..1747
 FT Domain /note= "thioesterase domain, TE7"
 PN EP-791656-A2.
 PD 27-AUG-1997.
 PF 19-FEB-1997; 301066.
 PR 22-FEB-1996; US-012050.
 PA (ELIL) LILLY & CO ELI.
 PI Burgett SQ, Kuhstoss SA, Rao RN, Richardson MA;
 PI Rosteck PR;
 DR WPI; 97-418047/39.
 DR N-PSDB; T78508.
 PT DNA encoding Streptomyces ambofaciens platenolide synthase domain -
 PT for production of spiramycin-related polyketide antibiotics
 PS Claim 8; Pages 71-77; 81pp; English.
 CC W23716-W23720 represent proteins encoded by the platenolide synthase gene
 CC cluster of the invention. The gene cluster is also referred to as the
 CC srmg gene, and was isolated from Streptomyces ambofaciens. These
 CC sequences are multi-functional proteins which direct the synthesis of the
 CC polyketide platenolide. Platenolide is the basic building block of the
 CC macrolide antibiotic spiramycin. The DNA can be used to produce compounds
 CC exhibiting antibiotic activity based on the platenolide structure,
 CC including specifically the macrolide antibiotic spiramycin and spiramycin
 CC analogues and derivatives. Modifications of the platenolide synthase DNA
 CC sequence can be made so as to change the number and type of carboxylic
 CC acids incorporated into the growing polyketide chain and to change the
 CC kind of post-condensation processing that is conducted.
 CC Sequence 1891 AA;

Query Match 76.8%; Score 43; DB 26; Length 1891;
 Best Local Similarity 75.0%; Pred. No. 2.04e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 1007 taglhryv 1014
 :|||||
 QY 57 SAGLHLRV 64

RESULT 15
 ID W77578 standard; Protein; 112 AA.
 AC W77578;
 DT 30-OCT-1998 (first entry)
 DE Staphylococcus aureus protein of unknown function.
 KW Staphylococcus aureus protein; immune response induction; eye infection;
 KW antibody production; T-cell immune response; gastrointestinal infection;
 KW respiratory infection; inhibitor; bacterial infection; cardiac infection;
 KW central nervous system; kidney infection; urinary tract infection;
 KW antimicrobial compound identification; broad spectrum antibiotic;
 KW therapy.
 OS Staphylococcus aureus.
 FH Key Location/Qualifiers
 FT Misc_difference 1..112
 FT /note= "residues designated X are unspecified, and
 represented as Xaa in the specification"
 PN EP-841394-A2.
 PD 13-MAY-1998.
 PF 24-SEP-1997; 307485.
 PR 24-SEP-1996; US-027032.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 PI Black MT, Burnham MKR, Hodgson JE, Knowles DJC,
 PI Lonetto MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M,
 PI Ward JM;
 DR WPI; 98-252940/23.
 DR N-PSDB; V53376.
 PT New nucleic acid sequences from Staphylococcus aureus WCHU29 -
 PT useful in vaccines and for treatment of bacterial infections of e.g.
 PT respiratory tract and central nervous system
 PS Claim 11; Page 263; 390pp; English.
 CC This sequence represents a Staphylococcus aureus protein of unknown
 CC function, and is encoded by a DNA sequence of the invention.
 CC The DNA sequences were isolated from Staphylococcus aureus WCHU29
 CC (NCIMB 40771). Host cells containing the DNA sequences are used to
 CC produce polypeptides or fragments. The proteins are used in the treatment
 CC of disease, for inducing an immune response by administering them, to
 CC produce antibody and/or T-cell immune response. Antagonists of the
 CC proteins are used for the inhibition of bacterial polypeptides.
 CC Conditions which may be treated include bacterial infections, especially
 CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,
 CC urinary tract, skin, bones and joints. The proteins can also be used to
 CC identify antimicrobial compounds which are broad spectrum antibiotics,
 CC especially useful in the treatment of H. pylori infection.
 CC Sequence 112 AA;

Query Match 75.0%; Score 42; DB 34; Length 112;
 Best Local Similarity 62.5%; Pred. No. 2.66e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 40 nsp1hryv 47
 :|||||
 QY 57 SAGLHLRV 64

Search completed: Wed Sep 1 16:28:39 1999
 Job time : 25 secs.

 W A S E N E

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed Sep 1 16:27:45 1999; MasPar time 3.59 Seconds
 Tabular output not generated.
 89.384 Million cell updates/sec

Title: >PCT-US99-13024-2
 Description: (57-64) from PCTUS9913024.pap (10 of 12)
 Perfect Score: 56
 Sequence: 1 SAGLHLRV 8

Scoring table: PAM 150
 Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: pir60
 1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 21.865; Variance 24.608; scale 0.889

pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	48	85.7	415	2	T02145	hypothetical protein
2	47	83.9	303	2	S10927	integrase-like protei
3	47	83.9	337	2	C26839	integrase - Escherich
4	47	83.9	337	2	I39499	integrase - Acinetoba
5	47	83.9	337	2	S32184	integrase - Klebsiell
6	47	83.9	337	2	A42646	integrase - Pseudomon
7	47	83.9	337	2	J00300	hypothetical 38.4K pr
8	45	80.4	499	2	F69833	monooxygenase homolog
9	44	78.6	147	2	D27733	hypothetical protein
10	44	78.6	461	2	S52408	SPDVR1 protein - sea
11	44	78.6	572	1	VEHULC	lamin C - human
12	44	78.6	574	2	S04333	lamin C - mouse
13	44	78.6	657	2	S05517	lamin A - human
14	44	78.6	664	1	VEHULA	lamin A - human
15	44	78.6	665	2	S28182	lamin A - mouse
16	44	78.6	665	2	S27267	lamin A - rat
17	43	76.8	341	1	H69401	probable radical-form
18	43	76.8	373	2	A27817	lignin peroxidase (EC
19	43	76.8	513	2	A33966	ntrA protein - Rhizob
20	43	76.8	525	2	A36130	transcription initiat
21	43	76.8	1117	2	S38673	desmoglein 2 - human
22	42	75.0	211	2	S65049	low molecular weight
23	42	75.0	218	2	T01412	low molecular weight

24	42	75.0	325	2	JN0148	neccin, brain - mouse	5.97e+01
25	42	75.0	349	2	E70640	hypothetical protein	5.97e+01
26	42	75.0	354	2	JE0309	ornithine carbamoyltr	5.97e+01
27	42	75.0	380	2	T01706	hypothetical protein	5.97e+01
28	42	75.0	461	2	S54543	probable membrane pro	5.97e+01
29	42	75.0	489	2	S69027	ammonium transport pr	5.97e+01
30	42	75.0	494	2	S62902	legumin 2 precursor -	5.97e+01
31	42	75.0	511	2	A40110	nicotinic acetylcholi	5.97e+01
32	42	75.0	524	2	F70579	probable cell divisio	5.97e+01
33	42	75.0	604	2	S25203	smr protein - Strept	5.97e+01
34	42	75.0	627	1	BHTLD	hemocyanin chain d -	5.97e+01
35	42	75.0	698	2	A47203	protein-glutamine gam	5.97e+01
36	42	75.0	702	2	C64835	hypothetical protein	5.97e+01
37	42	75.0	730	2	A70680	hypothetical protein	5.97e+01
38	42	75.0	818	2	JC4397	peroxinectin precurs	5.97e+01
39	42	75.0	862	2	D70199	DNA mismatch repair p	5.97e+01
40	42	75.0	1073	1	OYRHX	heat-stable enterotox	5.97e+01
41	42	75.0	1075	1	OYRTHX	heat-stable enterotox	5.97e+01
42	42	75.0	1287	2	A41685	SIL protein - human	5.97e+01
43	42	75.0	1706	2	T8499	zinc finger protein R	5.97e+01
44	41	73.2	509	3	T02942	O-succinylhomoserine	9.49e+01
45	41	73.2	5262	2	T03454	ALR protein - human	9.49e+01

ALIGNMENTS

RESULT 1
 ENTRY T02145 #type complete
 TITLE hypothetical protein F8K4.20 - Arabidopsis thaliana
 ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
 cress
 DATE 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change
 05-Mar-1999
 T02145
 Z14574
 Vysotskaya, V.S.; Schwartz, J.R.; Toriumi, M.; Kwan, A.; Yu,
 G.; Oji, O.; Liu, S.; Li, J.; Araujo, R.; Au, M.; Brendel,
 V.; Buehler, E.; Conway, A.B.; Conway, A.R.; Dewar, K.;
 Feng, J.; Kim, C.; Kurtz, D.; Li, Y.; Palm, C.J.; Shinn,
 P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel, N.A.;
 Theologis, A.
 #submission submitted to the EMBL Data Library, August 1998
 #description Arabidopsis thaliana chromosome 1 BAC F8K4 sequence.
 #accession T02145
 #status preliminary; translated from GB/EMBL/DBJ
 #molecule_type DNA
 #residues 1-415 #label VYS
 #cross-references EMBL:AC004392; NID:g3282170; PID:g3367531

GENETICS
 #map_position 1
 #note F8K4.20
 SUMMARY #length 415 #molecular-weight 46438 #checksum 9304
 Query Match 85.7%; Score 48; DB 2; Length 415;
 Best Local Similarity 100.0%; Pred. No. 3.06e+00;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 14 GLHLRV 19
 |||||
 QY 59 GLHLRV 64

RESULT 2
 ENTRY S10927 #type complete
 TITLE integrase-like protein 2M - Mycobacterium fortuitum
 ORGANISM #formal_name Mycobacterium fortuitum
 DATE 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change
 08-Sep-1997
 S10927
 S10927
 Martin, C.; Timm, J.; Rauszler, J.; Gomez-Lus, R.; Davies, J.;
 Gicquel, B.

```

#journal Nature (1990) 345:739-743
#title Transposition of an antibiotic resistance element in mycobacteria.
#cross-references EMBL:X3635; NID:g44283; PID:g44286
#accession S10927
#molecule_type DNA
#residues 1-303 #label NAT
#cross-references EMBL:X3635; NID:g44283; PID:g44286
CLASSIFICATION #superfamily integrase-like protein
SUMMARY #length 303 #molecular-weight 34884 #checksum 5690

Query Match 83.9%; Score 47; DB 2; Length 303;
Best Local Similarity 75.0%; Pred. NO. 5.12e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155
Qy 57 SAGLHLRV 64

RESULT 3
ENTRY C26839 #type complete
TITLE integrase - Escherichia coli plasmid and transposons
ALTERNATE_NAMES recombinase
ORGANISM #formal_name Escherichia coli
DATE 20-Jun-1989 #sequence_revision 20-Jun-1989 #text_change 13-Mar-1998
ACCESSIONS C26839; S21844; S31856; S04809; I60356; S60664
REFERENCE A26839
#authors Hall, R.M.; Vockler, C.
#journal Nucleic Acids Res. (1987) 15:7491-7501
#title The region of the IncN plasmid R46 coding for resistance to beta-lactam antibiotics, streptomycin/spectinomycin and sulphonamides is closely related to antibiotic resistance segments found in IncW plasmids and in Tn21-like transposons
#cross-references MUID:88015610
#accession C26839
#molecule_type DNA
#residues 1-337 #label HAL
#cross-references GB:X06046; NID:g42204; PID:g42207
#experimental_source plasmid R46
#genetics PLA
REFERENCE S21844
#authors Sundstroem, L.; Swedberg, G.; Skold, O.
#submission Submitted to the EMBL Data Library, March 1991
#description Characterization of the transposon Tn5086, and its site-specifically inserted gene dhfrVII, which encodes a new type of trimethoprim-resistant dihydrofolate reductase.
#accession S21844
#status preliminary
#molecule_type DNA
#residues 1-145 #label SUN
#cross-references EMBL:X58425; NID:g43090; PID:g43091
#experimental_source transposon Tn5086
#genetics TRA
REFERENCE S31854
#authors Heikkila, E.; Skurnik, M.; Sundstroem, L.; Huovinen, P.
#submission Submitted to the EMBL Data Library, February 1993
#description Another dihydrofolate reductase cassette, inserted in an integron, borne on a Tn21-like element.
#accession S31856
#status preliminary
#molecule_type DNA
#residues 1-55 #label HEI
#cross-references EMBL:221672; NID:g49350; PID:g49353
#experimental_source Tn21-like element
#genetics TLE
REFERENCE S04809
#authors Sundstroem, L.; Radstroem, P.; Swedberg, G.; Skold, O.
#journal Mol. Gen. Genet. (1988) 213:191-201
#title Site-specific recombination promotes linkage between trimethoprim- and sulfonamide resistance genes. Sequence

```

```

#cross-references dhfrv and sul1 and a recombination active locus of Tn21.
#accession MUID:89039710
#status translation not shown
#molecule_type DNA
#residues 1-81,'p',83,'p',85-337 #label SU2
#cross-references EMBL:X12868; NID:g45672; PID:g45673
#experimental_source plasmid pLMO20, Tn21
#genetics T21
REFERENCE I60356
#authors Radstrom, P.; Skold, O.; Swedberg, G.; Flensburg, J.; Roy, P.H.; Sundstrom, L.
#journal J. Bacteriol. (1994) 176:3257-3268
#title Transposon Tn5090 of plasmid R751, which carries an integron, is related to Tn7, Mu, and the retroelements.
#cross-references MUID:94252994
#accession I60356
#status preliminary; translated from GB/EMBL/DDBJ
#molecule_type DNA
#residues 1-145 #label RES
#cross-references EMBL:X58425; NID:g43090; PID:g43091
#experimental_source plasmid R751, Tn5090
#genetics T50
REFERENCE S60664
#authors Adrian, P.V.; Thomson, C.J.; Klugman, K.P.; Ames, S.G.
#submission submitted to the EMBL Data Library, August 1995
#description Nucleotide sequence of the cassette-borne type XIII dihydrofolate reductase from an Escherichia coli strain isolated in South Africa.
#accession S60664
#status preliminary
#molecule_type DNA
#residues 1-191 #label ADR
#cross-references EMBL:X50802; NID:g1841698; PID:g950043
#experimental_source strain isolated in South Africa
GENETICS PLA
#gene int
#genome int
GENETICS TRA
#gene int
#genome int
GENETICS TLE
#gene int
#genome int
GENETICS T21
#gene int
#genome int
GENETICS T50
#gene int
#genome int
CLASSIFICATION #superfamily integrase-like protein
SUMMARY #length 337 #molecular-weight 38381 #checksum 9946

Query Match 83.9%; Score 47; DB 2; Length 337;
Best Local Similarity 75.0%; Pred. NO. 5.12e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155
Qy 57 SAGLHLRV 64

RESULT 4
ENTRY I39499 #type complete
TITLE integrase - Acinetobacter baumannii
ORGANISM #formal_name Acinetobacter baumannii
DATE 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 16-Feb-1997
ACCESSIONS I39499
REFERENCE I39499
#authors Elisha, B.G.; Steyn, L.M.
#journal Curr. Microbiol. (1991) 22:259-263

```



```

#title      Cloning of AAC(3) and AAD(2') genes from Acinetobacter:
#accession  I39499
##status    preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues  1-337 ##label RES
##cross-references EMBL:U17586; NID:g596249; PID:g596251
GENETICS
#gene       tnpl
CLASSIFICATION
#superfamily integrase-like protein
SUMMARY     #length 337 #molecular-weight 38381 #checksum 9946

Query Match      83.9%; Score 47; DB 2; Length 337;
Best Local Similarity 75.0%; Pred. No. 5.12e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155
| ||:|
QY 57 SAGLHLRV 64

RESULT 5
ENTRY
TITLE      #type complete
#formal_name Klebsiella pneumoniae
ORGANISM   06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
DATE       08-Sep-1997
ACCESSIONS S32184
REFERENCE   S32177
#authors    Radstroem, P.; Sundstrom, L.; Swedberg, G.; Flensburg, J.; Skold, O.
#submission Submitted to the EMBL Data Library, March 1993
#description The integrons of R751 and Tn21 are transposons related to the
              retroelements and to Tn7.
#accession  S32184
#status     preliminary
##molecule_type DNA
##residues  1-337 ##label RAD
##cross-references EMBL:X72585; NID:g288626; PID:g288634
CLASSIFICATION
#superfamily integrase-like protein
SUMMARY     #length 337 #molecular-weight 38358 #checksum 180

Query Match      83.9%; Score 47; DB 2; Length 337;
Best Local Similarity 75.0%; Pred. No. 5.12e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155
| ||:|
QY 57 SAGLHLRV 64

RESULT 6
ENTRY
TITLE      #type complete
#formal_name Pseudomonas aeruginosa
ORGANISM   04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change
DATE       12-Apr-1995
ACCESSIONS A42646
REFERENCE   A42646
#authors    Bissonnette, L.; Roy, P.H.
#journal    J. Bacteriol. (1992) 174:1248-1257
#title      Characterization of in0 of Pseudomonas aeruginosa plasmid
              pvs1, an ancestor of integrons of multiresistance plasmids
              and transposons of gram-negative bacteria.
#cross-references MUID:92138617
#contents   plasmid pvs1
#accession  A42646
#status     preliminary
##molecule_type DNA
##residues  1-337 ##label BIS
##note       sequence extracted from NCBI backbone (NCBIN:79394,
              NCBP:79395)
CLASSIFICATION
#superfamily integrase-like protein

```

```

SUMMARY     #length 337 #molecular-weight 38381 #checksum 9946

Query Match      83.9%; Score 47; DB 2; Length 337;
Best Local Similarity 75.0%; Pred. No. 5.12e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155
| ||:|
QY 57 SAGLHLRV 64

RESULT 7
ENTRY
TITLE      #type complete
#formal_name Pseudomonas aeruginosa
ORGANISM   07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
DATE       05-Dec-1997
ACCESSIONS J00300
REFERENCE   J00299
#authors    Wohlleben, W.; Arnold, W.; Bissonnette, L.; Pelletier, A.;
              Tanquay, A.; Roy, P.H.; Gamboa, G.C.; Barry, G.F.; Aubert,
              E.; Davies, J.; Kagan, S.A.
#journal    Mol. Gen. Genet. (1989) 217:202-208
#title      On the evolution of Tn21-like multiresistance transposons:
              sequence analysis of the gene (aacC1) for gentamicin
              acetyltransferase-3-I(AAC(3)-I), another member of the
              Tn21-based expression cassette.
#cross-references MUID:89364699
#contents   Plasmid R1033
#accession  J00300
##molecule_type DNA
##residues  1-337 ##label WOH
##cross-references GB:X15852; NID:g45765; PID:g45766; GB:U12338;
              NID:g530803; PID:g530804
##note       this sequence is encoded by transposon Tn1696
CLASSIFICATION
#superfamily integrase-like protein
SUMMARY     #length 337 #molecular-weight 38372 #checksum 196

Query Match      83.9%; Score 47; DB 2; Length 337;
Best Local Similarity 75.0%; Pred. No. 5.12e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155
| ||:|
QY 57 SAGLHLRV 64

RESULT 8
ENTRY
TITLE      #type complete
#formal_name Bacillus subtilis
ORGANISM   05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
DATE       24-Sep-1998
ACCESSIONS F69833
REFERENCE   A69580
#authors    Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
              Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
              Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
              A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
              Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
              Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
              Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
              Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
              Fabret, C.; Ferrari, E.; Fouiger, D.; Fritz, C.; Fujita,
              M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
              S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
              Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
              C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
              Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
              Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
              Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
              Kurita, K.; Lapidus, A.; Lardinols, S.; Lauber, J.;
              Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;

```



```

CLASSIFICATION #superfamily cytoskeletal keratin
KEYWORDS alternative splicing; coiled coil; heptad repeat; membrane
protein; nuclear membrane; phosphoprotein; structural
protein

FEATURE
2-572 #product lamin C #status predicted #label MAT\
2-33 #domain head #label HED\
34-388 #domain rod #label ROD\
34-70 #region coil 1A\
81-218 #region coil 1B\
243-388 #region coil 2\
266 #region heptad change of phase\
330 #region heptad change of phase\
417-420 #domain tail #label END\
SUMMARY #length 572 #molecular-weight 6769

Query Match 78.6%; Score 44; DB 1; Length 572;
Best Local Similarity 62.5%; Pred. No. 2.29e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 56 NAGLRLRI 63
:||||:
Qy 57 SAGLHLRV 64

RESULT 12
ENTRY S04333 #type complete
TITLE lamin C - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change
17-Mar-1999
ACCESSIONS S04333; S03704; S65931
REFERENCE Riedel, W.; Werner, D.
#authors Biochim. Biophys. Acta (1989) 1008:119-122
#journal Nucleotide sequence of the full-length mouse lamin C cDNA and
#title its deduced amino-acid sequence.
#cross-references MUID:89247440
#accession S04333
##molecule_type mRNA
##residues 1-574 ##label RTE
##cross-references EMBL:X14170; NID:g52864; PID:g52865
REFERENCE S06662
#authors Weber, K.; Plessmann, U.; Traub, P.
#journal FEBS Lett. (1989) 257:411-414
#title Maturation of nuclear lamin A involves a specific
carboxy-terminal trimming, which removes the
polyisoprenylation site from the precursor; implications
for the structure of the nuclear lamina.
#cross-references MUID:90060368
#accession S09704
##molecule_type protein
##residues 521-574 ##label WEB
REFERENCE S65931
#authors Nakajima, N.; Abe, K.
#journal FEBS Lett. (1995) 365:108-114
#title Genomic structure of the mouse A-type lamin gene locus
encoding somatic and germ cell-specific lamins.
#cross-references MUID:95300954
#accession S65931
##status translation not shown
##molecule_type DNA
##residues 1-117 ##label NAK
##cross-references EMBL:D49732; NID:g1100724
CLASSIFICATION #superfamily cytoskeletal keratin
KEYWORDS alternative splicing; coiled coil; heptad repeat;
intermediate filament; nucleus; phosphoprotein

FEATURE
2-33 #domain head #status predicted #label HED\
34-388 #domain rod #status predicted #label ROD\
34-218 #region coiled coil 1\
243-388 #region coiled coil 2\

```

```

266 #region heptad change of phase\
325 #region stutter\
330 #region heptad change of phase\
389-574 #domain tail #status predicted #label END\
417-420 #region nuclear location signal
SUMMARY #length 574 #molecular-weight 65405 #checksum 7573

Query Match 78.6%; Score 44; DB 2; Length 574;
Best Local Similarity 62.5%; Pred. No. 2.29e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 56 NAGLRLRI 63
:||||:
Qy 57 SAGLHLRV 64

RESULT 13
ENTRY S05517 #type complete
TITLE lamin - chicken
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
08-Sep-1997
ACCESSIONS S05517
REFERENCE Peter, M.; Kitten, G.T.; Lehner, C.F.; Vorburger, K.; Bailler,
S.M.; Maridor, G.; Nigg, E.A.
#authors J. Mol. Biol. (1989) 208:393-404
#journal Cloning and sequencing of cDNA clones encoding chicken lamins
#title A and B(1) and comparison of the primary structures of
vertebrate A- and B-type lamins.
#cross-references MUID:90012208
#accession S05517
##status preliminary
##molecule_type mRNA
##residues 1-657 ##label PET
##cross-references EMBL:X16879; NID:g63555; PID:g63556
CLASSIFICATION #superfamily cytoskeletal keratin
KEYWORDS nucleus
SUMMARY #length 657 #molecular-weight 73164 #checksum 1468

Query Match 78.6%; Score 44; DB 2; Length 657;
Best Local Similarity 62.5%; Pred. No. 2.29e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 55 NAGLRLRI 62
:||||:
Qy 57 SAGLHLRV 64

RESULT 14
ENTRY VERULA #type complete
TITLE lamin A - human
ALTERNATE_NAMES 70KDa lamin
ORGANISM #formal_name Homo sapiens #common_name man
DATE 28-May-1986 #sequence_revision 04-Dec-1986 #text_change
05-Dec-1998
ACCESSIONS A02961; B24249; C24249
REFERENCE A02962
#authors Mckeen, F.D.; Kirschner, M.W.; Caput, D.
#journal Nature (1986) 319:463-468
#title Homologies in both primary and secondary structure between
nuclear envelope and intermediate filament proteins.
#cross-references MUID:86118697
#accession A02961
##molecule_type mRNA
##residues 1-582, 'LAHRAVRLRAACRQICQRLRSPGGRTHL',
'LWLFCLQCHGHSQLPQCGWGQWLRGSG',
'HPLPPGLOLPPNPPELQHHVWDLPGR',
'GGGGFLRPPLMPPCPARHGRLKAKK' ##label MCK
##cross-references GB:X03444; NID:g34227; PID:g34228
##note this sequence has been corrected in reference A94121; an
omitted nucleotide caused a reading frame error
beginning at position 583

```

REFERENCE A94121
#authors Fisher, D.Z.; Chaudhary, N.; Blobel, G.
#journal Proc. Natl. Acad. Sci. U.S.A. (1986) 83:6450-6454
#title cDNA sequencing of nuclear lamins A and C reveals primary and secondary structural homology to intermediate filament proteins.
#cross-references MUID:86313596
#note parts of sequences from rat lamins, but not human, were determined by protein sequencing
#accession B24249
##molecule_type mRNA
##residues 537-664 #label FIS
##note sequence fragment shown in publication
#accession C24249
##molecule_type mRNA
##residues 150-664 #label FI2
##cross-references GB:M13452; NID:9186838; PID:9386856
##note submitted sequence extracted from GenBank
#COMMENT Lamins A and C (see PIR:VEHULC) are products of alternative splicing of the same gene.
#COMMENT The lamins (A, B, and C) contains several alpha-helical domains capable of forming coiled coils.
#COMMENT The association of lamins, dependent upon ionic interactions, is interrupted by hyperphosphorylation during the interphase-metaphase period of the cell cycle. The nuclear envelope disintegrates with lamin dissociation; it does not reform until telophase, when the lamins are dephosphorylated to interphase levels and then reassociate.

GENETICS
#gene GDB:LMNA; LMN1
#cross-references GDB:I32146; OMIM:150330
#map_position lq21.2-lq21.3
#FUNCTION
#description structural component of the nuclear lamina, a fibrous meshwork on the nucleoplasmic surface of the nuclear membrane
#CLASSIFICATION superfamily cytoskeletal keratin
#KEYWORDS lipoprotein; membrane protein; methylated carboxyl end; nuclear membrane; phosphoprotein; prenylated cysteine; structural protein
#FEATURE
2-661 #product lamin A #status predicted #label MAT\
2-33 #domain head #label HED\
34-388 #domain rod #label ROD\
34-70 #region coil 1A\
81-218 #region coil 1B\
243-388 #region coil 2\
325 #region stutter\
389-661 #domain tail #label END\
417-420 #region nuclear location signal\
661 #binding_site farnesyl (Cys) (covalent) #status predicted
661 #modified_site methyl ester carboxyl end (Cys) (in mature form) #status predicted
#length 664 #molecular-weight 74139 #checksum 9955
#SUMMARY
Query Match 78.6%; Score 44; DB 1; Length 664;
Best Local Similarity 62.5%; Pred. No. 2-29e-01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 56 NAGLRLRI 63
:|||||:
QY 57 SAGLHLRV 64

RESULT 15
ENTRY S28182 #type complete
TITLE lamin A - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 17-Mar-1999
ACCESSIONS S28182; S06662; S65931

REFERENCE S28182
#authors Nakajima, N.; Sado, T.
#journal Blochim. Biophys. Acta (1993) 1171:311-314
#title Nucleotide sequence of a mouse lamin A cDNA and its deduced amino acid sequence.
#cross-references MUID:93144345
#accession S28182
##status preliminary
##molecule_type mRNA
##residues 1-665 #label NAK
#REFERENCE S06662
#authors Weber, K.; Plessmann, U.; Traub, P.
#journal FEBS Lett. (1989) 257:411-414
#title Maturation of nuclear lamin A involves a specific carboxy-terminal trimming, which removes the polyisoprenylation site from the precursor; implications for the structure of the nuclear lamina.
#cross-references MUID:90060368
#accession S06662
##molecule_type protein
##residues 520-580, 'X', 582-622, 'V', 624-647 #label WEB
#REFERENCE S65931
#authors Nakajima, N.; Abe, K.
#journal FEBS Lett. (1995) 365:108-114
#title Genomic structure of the mouse A-type lamin gene locus encoding somatic and germ cell-specific lamins.
#cross-references MUID:95300954
#accession S65931
##status translation not shown
##molecule_type DNA
##residues 1-117 #label NAW
#CLASSIFICATION superfamily cytoskeletal keratin
#KEYWORDS alternative splicing; coiled coil; heptad repeat; intermediate filament; lipoprotein; nucleus; phosphoprotein; prenylated cysteine
#FEATURE
1-33 #domain head #status predicted #label HED\
34-388 #domain rod #status predicted #label ROD\
34-218 #region coiled coil 1\
243-388 #region coiled coil 2\
266 #region heptad change of phase\
325 #region stutter\
330 #region heptad change of phase\
389-662 #domain tail #status predicted #label END\
417-420 #region nuclear location signal\
662 #binding_site farnesyl (Cys) (covalent) #status predicted
#SUMMARY #length 665 #molecular-weight 74184 #checksum 676
#Query Match 78.6%; Score 44; DB 2; Length 665;
#Best Local Similarity 62.5%; Pred. No. 2-29e-01;
#Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 56 NAGLRLRI 63
:|||||:
QY 57 SAGLHLRV 64

Search completed: Wed Sep 1 16:27:57 1999
Job time : 12 secs.

WIREH

(TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:26:37 1999; MasPar time 2.52 Seconds
Tabular output not generated. 89.698 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (57-64) from PCTUS9913024.pep (10 of 12)
Perfect Score: 56
Sequence: 1 SAGLHLRV 8

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 22.517; Variance 22.845; scale 0.986

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description	Pred. No.
1	47	83.9	1	INTR_ECOLI	2.31e+00
2	44	78.6	1	RBL_PTEVI	1.16e+01
3	44	78.6	1	DVR1_STRPU	1.16e+01
4	44	78.6	1	LAMC_HUMAN	1.16e+01
5	44	78.6	1	LAMC_MOUSE	1.16e+01
6	44	78.6	1	LAMA_CHICK	1.16e+01
7	44	78.6	1	LAMA_HUMAN	1.16e+01
8	44	78.6	1	LAMA_MOUSE	1.16e+01
9	44	78.6	1	LAMA_RAT	1.16e+01
10	43	76.8	1	PHTE_PSEPU	1.94e+01
11	43	76.8	1	RP55_RHIME	1.94e+01
12	43	76.8	1	RP54_RHIME	1.94e+01
13	43	76.8	1	RP54_RHISN	1.94e+01
14	43	76.8	1	DESMOGLIN 2	1.94e+01
15	43	76.8	1	VIT6_OSCBR	1.94e+01
16	42	75.0	1	HS2M_SOYBN	3.23e+01
17	42	75.0	1	NECD_MOUSE	3.23e+01
18	42	75.0	1	MEP3_YEAST	3.23e+01
19	42	75.0	1	ACH2_RAT	3.23e+01
20	42	75.0	1	NEURONAL ACETYLCHOLINE	3.23e+01
21	42	75.0	1	CYTCHROME P450 4E5 PR	3.23e+01
22	42	75.0	1	HYPOHETICAL 56.3 KD F	3.23e+01
23	42	75.0	1	HOCVD_EURCA	3.23e+01
24	42	75.0	1	TGLC_CHICK	3.23e+01

24 42 75.0 702 1 YCBY_ECOLI HYPOTHETICAL 78.9 KD P 3.23e+01
25 42 75.0 862 1 MUTS_BORBU DNA MISMATCH REPAIR PR 3.23e+01
26 42 75.0 1072 1 HSER_RAT HEAT-STABLE ENTEROTOXI 3.23e+01
27 42 75.0 1073 1 HSER_PIG HEAT-STABLE ENTEROTOXI 3.23e+01
28 42 75.0 1073 1 HSER_HUMAN HEAT-STABLE ENTEROTOXI 3.23e+01
29 42 75.0 1076 1 HSER_CAVPO HEAT-STABLE ENTEROTOXI 3.23e+01
30 41 73.2 72 1 YCA1_OENBE HYPOTHETICAL 8.5 KD PR 5.31e+01
31 41 73.2 149 1 Y213_METJA HYPOTHETICAL PROTEIN M 5.31e+01
32 41 73.2 214 1 RL10_CAEEL 60S RIBOSOMAL PROTEIN 5.31e+01
33 41 73.2 228 1 RL10_PINTA 60S RIBOSOMAL PROTEIN 5.31e+01
34 41 73.2 247 1 ZO20_XENLA OOCYTE ZINC FINGER PRO 5.31e+01
35 41 73.2 328 1 Y052_HAEIN HYPOTHETICAL PROTEIN H 5.31e+01
36 41 73.2 340 1 Y0BN_MYCTU HYPOTHETICAL 37.2 KD P 5.31e+01
37 41 73.2 364 1 F16P_XANFL FRUCTOSE-1,6-BISPHOSPH 5.31e+01
38 41 73.2 426 1 SLS1_YARLI SLS1 PROTEIN PRECURSOR 5.31e+01
39 41 73.2 433 1 DHOM_SYNY3 HOMOSERINE DEHYDROGENA 5.31e+01
40 41 73.2 563 1 METB_ARATH CYSTATHIONINE GAMMA-SY 5.31e+01
41 41 73.2 611 1 YX45_MYCTU HYPOTHETICAL 67.2 KD P 5.31e+01
42 41 73.2 1130 1 YL17_CAEEL HYPOTHETICAL 131.5 KD 5.31e+01
43 41 73.2 1308 1 YTX2_XENLA TRANSPONSON TX1 HYPOTHE 5.31e+01
44 41 73.2 1547 1 RRPQ_PMV RNA REPLICATION PROTEI 5.31e+01
45 41 73.2 1841 1 SUI5_RAT SUCRASE-ISOMALTASE, IN 5.31e+01

ALIGNMENTS

RESULT 1
ID INTR_ECOLI STANDARD; PRT; 337 AA.
AC P09999; P11743;
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE INTEGRASE/RECOMBINASE (E2 PROTEIN).
GN INT.
OS ESCHERICHIA COLI, AND PSEUDOMONAS AERUGINOSA.
OG PLASMID INCN R46, PLASMID RGN238, PLASMID RI033, PLASMID PLMO20,
OG PLASMID PLMO27, PLASMID PLMO150, PLASMID PLMO229, AND PLASMID PVS1.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC PLASMID-INCN R46;
RX MEDLINE; 88015610.
RA HALL R.M., VOCKLER C.;
RT "The region of the IncN plasmid R46 coding for resistance to
beta-lactam antibiotics, streptomycin/spectinomycin and sulphonamides
is closely related to antibiotic resistance segments found in IncW
plasmids and in Tn21-like transposons.";
RT NUCLEIC ACIDS RES. 15:7491-7501(1987).
RN [2]
RP PARTIAL SEQUENCE FROM N.A.
RC PLASMID-RGN238;
RX MEDLINE; 88041110.
RA OUELLETTE M., BISSONNETTE L., ROY P.H.;
RT "Precise insertion of antibiotic resistance determinants into
Tn21-like transposons: nucleotide sequence of the OXA-1
beta-lactamase gene.";
RT PROC. NATL. ACAD. SCI. U.S.A. 84:7378-7382(1987).
RN [3]
RP SEQUENCE FROM N.A.
RC PLASMID-RGN238;
RX OUELLETTE M.;
RT SUBMITTED (DEC-1987) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RC PLASMID-PLMO20;
RX MEDLINE; 89039710.
RA SUNDTSTROM L., RADSTROM P., SWEDBERG G., SKOELD O.;
RT "Site-specific recombination promotes linkage between trimethoprim-
and sulfonamide resistance genes. Sequence characterization of dhfrv
and sulI and a recombination active locus of Tn21.";
RL MOL. GEN. GENET. 213:191-201(1988).
RN [5]

RP SEQUENCE FROM N.A.
 RC SPECIES-P.AERUGINOSA; PLASMID=RI033; TRANSPOSON-TN1696;
 RX MEDLINE: 89364699.
 RA WOHLLEBEN W., ARNOLD W., BISSONNETTE L., PALLETIER A., TANGUAY A.,
 ROY P.H., CAMBOA G.C., BARRY G.F., AUBERT E., DAVIES J., KAGAN S.A.;
 RT "On the evolution of *tn21*-like multiresistance transposons: sequence
 analysis of the gene (aacC1) for gentamicin acetyltransferase-3-
 I(AAC(3)-I), another member of the *Tn21*-based expression cassette.";
 RL MOL. GEN. GENET. 217:202-208(1989).
 RN [6]
 RP SEQUENCE OF 1-145 FROM N.A.
 RC PLASMID-PLM0150, AND PLM0229;
 RX MEDLINE: 90262183.
 RA SUNDTROEM L., SKOELD O.;
 RT "The dhfr1 trimethoprim resistance gene of *Tn7* can be found at
 specific sites in other genetic surroundings.";
 RL ANTIMICROB. AGENTS CHEMOTHER. 34:642-650(1990).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC SPECIES-P.AERUGINOSA; PLASMID=PVSL1;
 RX MEDLINE: 92138617.
 RA BISSONNETTE L., ROY P.H.;
 RT "Characterization of *In0* of *Pseudomonas aeruginosa* plasmid pVSL1, an
 ancestor of integrons of multiresistance plasmids and transposons of
 gram-negative bacteria.";
 RL J. BACTERIOL. 174:1248-1257(1992).
 RN [8]
 RP SEQUENCE OF 1-145 FROM N.A.
 RC SPECIES-E.COLI; PLASMID-PLM027; TRANSPOSON-TN5086;
 RA SUNDTROEM L., SWEDBERG G., SKOLD O.;
 RL SUBMITTED (MAR-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -!- FUNCTION: PUTATIVE INTEGRASE BELIEVED TO BE INVOLVED IN
 INSERTIONS OF ANTIBIOTIC RESISTANCE GENES INTO PLASMIDS
 AND TRANSPOSONS.
 CC -!- THE SEQUENCE SHOWN IS THAT OF PLASMID R46.
 CC -!- SIMILARITY: BELONGS TO THE "PHAGE" INTEGRASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U12338; G530804;
 DR EMBL: U49101; G530814;
 DR EMBL: X06046; G42207;
 DR EMBL: M95287; G151817;
 DR EMBL: J02957; G152063;
 DR EMBL: X12868; G45673;
 DR EMBL: X15852; G45766;
 DR EMBL: X12870; G48206;
 DR EMBL: X17478; G45584;
 DR EMBL: X17477; G45579;
 DR EMBL: M73819; G151300;
 DR EMBL: X58425; G43091;
 DR PIR: C26839; C26839.
 DR PIR: A42646; A42646.
 DR PFAM: PF00589; Phage_integrase; 1.
 DR HSSP: P21891; LAOP.
 KW DNA RECOMBINATION; DNA INTEGRATION; PLASMID; TRANSPOSABLE ELEMENT.
 FT VARIANT 32 32 R -> P (IN PLM0150).
 FT VARIANT 39 39 H -> N (IN R1033).
 FT VARIANT 82 82 A -> P (IN PLM020).
 FT VARIANT 84 84 A -> P (IN PLM020).
 FT VARIANT 187 187 S -> T (IN R1033).
 FT VARIANT 205 205 E -> Q (IN RGN238).
 FT ACT_SITE 312 312 PROBABLE TRANSIENT COVALENT LINKAGE TO
 DNA DURING STRAND CLEAVAGE AND REJOINING.
 FT SEQUENCE 337 AA; 38381 MW; C7BE628D CRC32;
 Query Match 83.9%; Score 47; DB 1; Length 337;

Best Local Similarity 75.0%; Pred. No. 2.31e+00;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Db 148 SEGQLRV 155
 QY 57 SAGHLRV 64
 I:|:|:|
 I:|:|:|
 RESULT 2
 ID RBL_PTEVI STANDARD; PRT; 440 AA.
 AC Q33015;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN (EC 4.1.1.39) (RUBISCO
 DE LARGE SUBUNIT) (FRAGMENT).
 GN RBCL.
 GN RBCL.
 OS PTERIS VITTATA (CHINESE BRAKE).
 OS CHLOROPLAST.
 OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLOPHYTES; FILICOPHYTA; FILICOPSIDA; FILICALES; PTERIDACEAE;
 OC PTERIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-LEAF;
 RX MEDLINE: 95211278.
 RA WOLF P.G., SOLTIS P.S., SOLTIS D.E.;
 RT "Phylogenetic relationships of demstaedtioid ferns: evidence from
 rbcL sequences.";
 RL MOL. PHYLOGENET. EVOL. 3:383-392(1994).
 CC -!- FUNCTION: RUBISCO CATALYSES TWO REACTIONS: THE CARBOXYLATION OF
 D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 ACTIVE SITE.
 CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) -
 2 3-PHOSPHO-D-GLYCERATE.
 CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) -
 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCULATE.
 CC -!- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
 CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.
 CC -!- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U05941; G463542;
 DR PROSITE: P500157; RUBISCO_LARGE; 1.
 DR PFAM: PF00016; RUBISCO_large; 1.
 DR MENDEL: 3802; PTEVI;rbcl;1.
 DR HSSP: P00875; LAUS.
 KW PHOTOSYNTHESIS; CARBON DIOXIDE FIXATION; PHOTORESPIRATION;
 KW LIASE; OXIDOREDUCTASE; MONOOXYGENASE; CHLOROPLAST.
 FT NON_TER 1 1
 FT ACT_SITE 191 191
 FT NON_TER 440 440
 FT SEQUENCE 440 AA; 48572 MW; BDC52A9 CRC32;
 Query Match 78.6%; Score 44; DB 1; Length 440;
 Best Local Similarity 83.3%; Pred. No. 1.16e+01;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 298 GMLRV 303
 QY 59 GLHLRV 64
 I:|:|:|
 I:|:|:|

```

RESULT 3
ID DVL1_STRPU STANDARD; PRT; 461 AA.
AC P48969;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE DVL1 PROTEIN HOMOLOG PRECURSOR.
GN DVL1
OS STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC ECHINOIDEA; ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE;
OC STRONGYLOCENTROTUS.
RN [1]
RP SEQUENCE FROM N.A.
RA PONCE M.R., MICOL J.L., DAVIDSON E.H.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z48313; G673497; -
CC PROSITE; PS00250; TGF-BETA; 1.
CC PFAM; PF00019; TGF-beta; 1.
CC HSP; P18075; IBMP.
CC GROWTH FACTOR; CYTOKINE; GLYCOPROTEIN; SIGNAL.
CC SIGNAL 1 30 POTENTIAL.
CC PROPEP 31 338 POTENTIAL.
CC CHAIN 339 461 DVR-1 PROTEIN HOMOLOG.
CC DISULFID 360 426 BY SIMILARITY.
CC DISULFID 389 458 BY SIMILARITY.
CC DISULFID 393 460 BY SIMILARITY.
CC DISULFID 425 425 INTERCHAIN (BY SIMILARITY).
CC CARBOHYD 149 149 POTENTIAL.
CC CARBOHYD 402 402 POTENTIAL.
CC SEQUENCE 461 AA; 51881 MW; 57C7CBA0 CRC32;

Query Match 78.6%; Score 44; DB 1; Length 461;
Best Local Similarity 62.5%; Pred. No. 1.16e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 267 NVGLQLEV 274
QY 57 SAGLHLRV 64

RESULT 4
ID LAMC_HUMAN STANDARD; PRT; 572 AA.
AC P02546;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE LAMIN C.
GN LAMIN C.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA MCKEON F.D., KIRSCHNER M.W., CAPUT D.;
RL "Homologies in both primary and secondary structure between nuclear
RL envelope and intermediate filament proteins."
RL NATURE 319:463-468(1986).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE; 86313596.
RA FISHER D.Z., CHAUDHARY N., BLOBEL G.;
RT "cDNA sequencing of nuclear lamins A and C reveals primary and
RT secondary structural homology to intermediate filament proteins.";
RL PROC. NATL. ACAD. SCI. U.S.A. 83:6450-6454(1986).
CC -1- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA. A FIBROUS
CC LAYER ON THE NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,
CC WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE
CC AND MAY ALSO INTERACT WITH CHROMATIN.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- PTM: INCREASED PHOSPHORYLATION OF THE LAMINS OCCURS BEFORE
CC ENVELOPE DISINTEGRATION AND PROBABLY PLAYS A ROLE IN REGULATING
CC LAMIN ASSOCIATIONS.
CC -1- THERE ARE THREE TYPES OF LAMINS IN HUMAN CELLS: A, B, AND C.
CC -1- THE STRUCTURAL INTEGRITY OF THE LAMINA IS STRICTLY CONTROLLED BY
CC THE CELL CYCLE, AS SEEN BY THE DISINTEGRATION AND FORMATION OF
CC THE NUCLEAR ENVELOPE IN PROPHASE AND TELOPHASE, RESPECTIVELY.
CC -1- ALTERNATIVE PRODUCTS: LAMINS A AND C ARE THE PRODUCTS OF ALTERNATE
CC SPLICING OF THE SAME GENE. THEY ARE PRESENT IN EQUAL AMOUNTS IN
CC THE LAMINA OF MAMMALS.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M13451; G307108; -
CC EMBL; X03445; G34236; -
CC PIR; A02962; VEHULC.
CC PIR; A24249; A24249.
CC MIM; 150330; -
CC PROSITE; PS00226; IF; 1.
CC PFAM; PF00038; filament; 1.
CC PFAM; PF00932; IF-tail; 1.
CC INTERMEDIATE FILAMENT; HEPTAD REPEAT PATTERN; COILED COIL;
CC NUCLEAR PROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING.
CC FT DOMAIN 1 33 HEAD.
CC FT DOMAIN 34 383 ROD.
CC FT DOMAIN 384 572 TAIL.
CC FT DOMAIN 71 70 COIL 1A.
CC FT DOMAIN 81 218 COIL 1B.
CC FT DOMAIN 219 242 LINKER 1.
CC FT DOMAIN 243 383 LINKER 2.
CC FT SITE 266 266 STUTTER (BY SIMILARITY).
CC FT SITE 330 330 HEPTAD CHANGE OF PHASE.
CC FT SITE 417 422 HEPTAD CHANGE OF PHASE.
CC FT DOMAIN 417 422 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC SEQUENCE 572 AA; 65134 MW; 6F6A20D0 CRC32;

Query Match 78.6%; Score 44; DB 1; Length 572;
Best Local Similarity 62.5%; Pred. No. 1.16e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 56 NAGLRLRI 53
QY 57 SAGLHLRV 64

RESULT 5
ID LAMC_MOUSE STANDARD; PRT; 574 AA.
AC P11516;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE LAMINS C AND C2.
GN LMN1 OR LMNA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

```

CC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
 RP [1]
 RX SEQUENCE FROM N.A.
 RA MEDLINE; 89247440.
 RA RIEDEL W., WERNER D.;
 RT "Nucleotide sequence of the full-length mouse lamin C cDNA and its
 RL deduced amino-acid sequence.";
 RL BICCHIM. BIOPHYS. ACTA 1008:119-122(1989).
 RN [2]
 RP SEQUENCE FROM N.A. (LAMIN C2).
 RC STRAIN-DDY; TISSUE-TESTIS;
 RX MEDLINE; 94244715;
 RA FURUKAWA K., INAGAKI H., HOTTA Y.;
 RT "Identification and cloning of an mRNA coding for a germ
 RL cell-specific A-type lamin in mice.";
 RL EXP. CELL RES. 212:426-430(1994).
 RN [3]
 RP SEQUENCE OF 235-568 FROM N.A.
 RX MEDLINE; 93144345.
 RA NAKAJIMA N., SADO T.;
 RT "Nucleotide sequence of a mouse lamin A cDNA and its deduced amino
 RL acid sequence.";
 RL BICCHIM. BIOPHYS. ACTA 1171:311-314(1993).
 RN [4]
 RP SEQUENCE OF 1-119 FROM N.A.
 RX MEDLINE; 95300954.
 RA NAKAJIMA N., ABE K.;
 RT "Genomic structure of the mouse A-type lamin gene locus encoding
 RL somatic and germ cell-specific lamins.";
 RL FEBS LETT. 365:108-114(1995).
 RN [5]
 RP PARTIAL SEQUENCE, AND PHOSPHORYLATION.
 RX MEDLINE; 92070490.
 RA EGGERT M., RADOMSKI N., TRIPIER D., TRAUB P., JOST E.;
 RT "Identification of phosphorylation sites on murine nuclear lamin C by
 RL RP-HPLC and microsequencing.";
 RL FEBS LETT. 292:205-209(1991).
 CC [1]- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS
 CC LAYER ON THE NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,
 CC WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE
 CC AND MAY ALSO INTERACT WITH CHROMATIN.
 CC [2]- SUBCELLULAR LOCATION: NUCLEAR.
 CC [3]- TISSUE SPECIFICITY: LAMIN C2 IS SPECIFICALLY EXPRESSED IN GERM
 CC CELLS. THIS SPECIFIC EXPRESSION AND UNIQUE STRUCTURE SUGGESTS A
 CC ROLE FOR LAMIN C2 IN DETERMINING THE ORGANIZATION OF NUCLEAR AND
 CC CHROMOSOMAL STRUCTURES DURING SPERMATOGENESIS.
 CC [4]- PTM: INCREASED PHOSPHORYLATION OF THE LAMINS OCCURS BEFORE
 CC ENVELOPE DISINTEGRATION AND PROBABLY PLAYS A ROLE IN REGULATING
 CC LAMIN ASSOCIATIONS.
 CC [5]- THE STRUCTURAL INTEGRITY OF THE LAMINA IS STRICTLY CONTROLLED BY
 CC THE CELL CYCLE, AS SEEN BY THE DISINTEGRATION AND FORMATION OF
 CC THE NUCLEAR ENVELOPE IN PROPHASE AND TELOPHASE, RESPECTIVELY.
 CC [6]- ALTERNATIVE PRODUCTS: LAMINS A, C AND C2 ARE THE PRODUCTS OF
 CC ALTERNATE SPLICING OF THE SAME GENE. LAMINS A AND C ARE PRESENT IN
 CC EQUAL AMOUNTS IN THE LAMINA OF MAMMALS.
 CC [7]- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X14170; G52865;
 CC EMBL; D14850; G505027;
 CC EMBL; D49732; E212319;
 CC EMBL; D49733; G1794160;
 CC EMBL; D49733; G1794161;
 CC EMBL; S04333; S04333;
 CC PIR; S18324; S18324;
 CC MGD; MGI:96794; LMNA.

DR PROSITE; PS00226; IF; 1.
 DR PFAM; PF00038; filament; 1.
 DR PFAM; PF00932; IF tail; 1.
 KW INTERMEDIATE FILAMENT; HEPTAD REPEAT PATTERN; COILED COIL;
 KW NUCLEAR PROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING.
 FT DOMAIN 1 33 HEAD.
 FT DOMAIN 34 383 ROD.
 FT DOMAIN 384 574 TAIL.
 FT DOMAIN 34 70 COIL 1A.
 FT DOMAIN 71 80 COIL 1B.
 FT DOMAIN 81 218 LINKER 1.
 FT DOMAIN 219 242 LINKER 2.
 FT DOMAIN 243 383 COIL 2.
 FT SITE 325 325 STUTTER (BY SIMILARITY).
 FT SITE 266 266 HEPTAD CHANGE OF PHASE.
 FT SITE 330 330 HEPTAD CHANGE OF PHASE.
 FT DOMAIN 417 422 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT MOD_RES 392 392 PHOSPHORYLATION (BY CDC2) (PARTIAL,
 FT MOD_RES 407 407 AT INTERPHASE).
 FT MOD_RES 409 409 PHOSPHORYLATION (PARTIAL, AT INTERPHASE).
 FT VARSPLIC 1 112 MISSING (IN LAMIN C2).
 FT VARSPLIC 113 118 FKELKA -> MGNAEG (IN LAMIN C2).
 FT CONFLICT 118 119 AR -> VC (IN REF. 1).
 FT CONFLICT 401 401 R -> P (IN REF. 1).
 FT CONFLICT 439 440 RV -> WL (IN REF. 1).
 SQ SEQUENCE 574 AA; 65446 MW; 3C8AD85A CRC32;
 Query Match 78.6%; Score 44; DB 1; Length 574;
 Best Local Similarity 62.5%; Pred. No. 1.16e+01;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Db 56 NAGLRLRI 63
 Qy 57 SAGLRLRV 64
 RESULT 6
 ID LAMA-CHICK STANDARD; PRT; 657 AA.
 AC P13648;
 DT 01-JAN-1990 (REL. 13, CREATED)
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
 DE LAMIN A.
 OS GALLUS GALLUS (CHICKEN).
 OC EUKARYOTA; METAZOA; VERTEBRATA; ARCHOSAURIA; AVES;
 OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; GALLUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 90012208.
 RA PETER M., KITTEL G.T., LEHNER C.F., VORBURGER K., BAILER S.M.,
 RA MARIDOR G., NIGG E.A.;
 RT "Cloning and sequencing of cDNA clones encoding chicken lamins A and
 RT B1 and comparison of the primary structures of vertebrate A- and
 RT B-type lamins.";
 RL J. MOL. BIOL. 208:393-404(1989).
 CC [1]- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS
 CC LAYER ON THE NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,
 CC WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE
 CC AND MAY ALSO INTERACT WITH CHROMATIN.
 CC [2]- SUBCELLULAR LOCATION: NUCLEAR.
 CC [3]- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X16879; G63556;
 CC PIR; S05517; S05517.

DR PROSITE; PS00226; IF; 1.
DR PFAM; PF00038; filament; 1.
DR PFAM; PF00932; IF tail; 1.
KW INTERMEDIATE FILAMENT; HEPTAD REPEAT PATTERN; COILED COIL;
KW NUCLEAR PROTEIN; LIPOPROTEIN; PRENYLATION; PHOSPHORYLATION;
KW ALTERNATIVE SPLICING.
FT DOMAIN 1 32 HEAD.
FT DOMAIN 33 382 ROD.
FT DOMAIN 383 657 TAIL.
FT DOMAIN 33 69 COIL 1A.
FT DOMAIN 70 79 LINKER 1.
FT DOMAIN 80 217 LINKER 2.
FT DOMAIN 218 241 COIL 1B.
FT DOMAIN 242 382 COIL 2.
FT LIPID 654 654 FARNESYL (BY SIMILARITY).
SQ SEQUENCE 657 AA; 73164 MW; 3E7C5858 CRC32;

Query Match 78.6%; Score 44; DB 1; Length 657;
Best Local Similarity 62.5%; Pred. No. 1.16e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 55 NAGLRLRI 62
QY 57 SAGLHLRV 64

RESULT 7
ID LAMA HUMAN STANDARD; PRT; 664 AA.

AC P02345; 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 04, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE LAMIN A (70 KD LAMIN).
GN LMN1 OR LMNA.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86118697.

RA MCKEON F.D.; KIRSCHNER M.W.; CAPUT D.;
RT "Homologies in both primary and secondary structure between nuclear envelope and intermediate filament proteins.";
RL NATURE 319:463-468(1986).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 583-644.
RX MEDLINE; 86313596.

RA FISHER D.Z.; CHAUDHARY N.; BLOBEL G.;
RT "CDNA sequencing of nuclear lamins A and C reveals primary and secondary structural homology to intermediate filament proteins.";
RL PROC. NATL. ACAD. SCI. U.S.A. 83:6450-6454(1986).
CC -!- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS LAYER ON THE NUCLEOLAR PLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE, AND MAY ALSO INTERACT WITH CHROMATIN.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- PTM: INCREASED PHOSPHORYLATION OF THE LAMINS OCCURS BEFORE LAMIN ASSOCIATIONS.
CC -!- THERE ARE THREE TYPES OF LAMINS IN HUMAN CELLS: A, B, AND C.

CC -!- THE STRUCTURAL INTEGRITY OF THE LAMINA IS STRICTLY CONTROLLED BY THE CELL CYCLE, AS SEEN BY THE DISINTEGRATION AND FORMATION OF THE NUCLEAR ENVELOPE IN PROPHASE AND TELOPHASE, RESPECTIVELY.
CC -!- ALTERNATIVE PRODUCTS: LAMINS A AND C ARE THE PRODUCTS OF ALTERNATE SPLICING OF THE SAME GENE. THEY ARE PRESENT IN EQUAL AMOUNTS IN THE LAMINA OF MAMMALS.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 582 ONWARD AND IS LONGER (702 AA) DUE TO A FRAMESHIFT.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----

CC EMBL; X03444; G34228; ALT_FRAME.
CC EMBL; M13452; G386856;
CC PIR; A02961; VEHULA.
CC PIR; B24249; B24249.
CC MIM; 150330;
CC DR PROSITE; PS00226; IF; 1.
CC DR PFAM; PF00038; filament; 1.
CC DR PFAM; PF00932; IF tail; 1.
KW INTERMEDIATE FILAMENT; HEPTAD REPEAT PATTERN; COILED COIL;
KW NUCLEAR PROTEIN; LIPOPROTEIN; PRENYLATION; PHOSPHORYLATION;
KW ALTERNATIVE SPLICING.

FT MOD_RES 1 1 BLOCKED.
FT DOMAIN 1 33 HEAD.
FT DOMAIN 34 383 ROD.
FT DOMAIN 384 664 TAIL.
FT DOMAIN 34 70 COIL 1A.
FT DOMAIN 81 218 LINKER 1.
FT DOMAIN 219 242 COIL 1B.
FT DOMAIN 243 383 LINKER 2.
FT SITE 325 325 STUTTER (BY SIMILARITY).
FT DOMAIN 417 422 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT LIPID 661 661 FARNESYL (BY SIMILARITY).
SQ SEQUENCE 664 AA; 74139 MW; 31F253EE CRC32;

Query Match 78.6%; Score 44; DB 1; Length 664;
Best Local Similarity 62.5%; Pred. No. 1.16e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 56 NAGLRLRI 63
QY 57 SAGLHLRV 64

RESULT 8
ID LAMA_MOUSE STANDARD; PRT; 665 AA.

AC P48678; P97859;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE LAMIN A.
GN LMN1 OR LMNA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95300954.
RA NAKAJIMA N.; ABE K.;
RT "Genomic structure of the mouse A-type lamin gene locus encoding somatic and germ cell-specific lamins.";
RL FEBS LETT. 365:108-114(1995).
RN [2]
RP SEQUENCE OF 1-568 FROM N.A.
RX MEDLINE; 89247440.
RA RIEDEL W.; WERNER D.;
RT "Nucleotide sequence of the full-length mouse lamin C cDNA and its deduced amino-acid sequence.";
RL BIOCHIM. BIOPHYS. ACTA 1008:119-122(1989).
RN [3]
RP SEQUENCE OF 235-665 FROM N.A.
RX MEDLINE; 93144345.
RA NAKAJIMA N.; SADO T.;
RT "Nucleotide sequence of a mouse lamin A cDNA and its deduced amino acid sequence.";
RL BIOCHIM. BIOPHYS. ACTA 1171:311-314(1993).
CC -!- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS LAYER ON THE NUCLEOLAR PLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,

CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----

CC EMBL; X03444; G34228; ALT_FRAME.
CC EMBL; M13452; G386856;
CC PIR; A02961; VEHULA.
CC PIR; B24249; B24249.
CC MIM; 150330;
CC DR PROSITE; PS00226; IF; 1.
CC DR PFAM; PF00038; filament; 1.
CC DR PFAM; PF00932; IF tail; 1.
KW INTERMEDIATE FILAMENT; HEPTAD REPEAT PATTERN; COILED COIL;
KW NUCLEAR PROTEIN; LIPOPROTEIN; PRENYLATION; PHOSPHORYLATION;
KW ALTERNATIVE SPLICING.

FT MOD_RES 1 1 BLOCKED.
FT DOMAIN 1 33 HEAD.
FT DOMAIN 34 383 ROD.
FT DOMAIN 384 664 TAIL.
FT DOMAIN 34 70 COIL 1A.
FT DOMAIN 81 218 LINKER 1.
FT DOMAIN 219 242 COIL 1B.
FT DOMAIN 243 383 LINKER 2.
FT SITE 325 325 STUTTER (BY SIMILARITY).
FT DOMAIN 417 422 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT LIPID 661 661 FARNESYL (BY SIMILARITY).
SQ SEQUENCE 664 AA; 74139 MW; 31F253EE CRC32;

Query Match 78.6%; Score 44; DB 1; Length 664;
Best Local Similarity 62.5%; Pred. No. 1.16e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 56 NAGLRLRI 63
QY 57 SAGLHLRV 64

RESULT 8
ID LAMA_MOUSE STANDARD; PRT; 665 AA.

AC P48678; P97859;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE LAMIN A.
GN LMN1 OR LMNA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95300954.
RA NAKAJIMA N.; ABE K.;
RT "Genomic structure of the mouse A-type lamin gene locus encoding somatic and germ cell-specific lamins.";
RL FEBS LETT. 365:108-114(1995).
RN [2]
RP SEQUENCE OF 1-568 FROM N.A.
RX MEDLINE; 89247440.
RA RIEDEL W.; WERNER D.;
RT "Nucleotide sequence of the full-length mouse lamin C cDNA and its deduced amino-acid sequence.";
RL BIOCHIM. BIOPHYS. ACTA 1008:119-122(1989).
RN [3]
RP SEQUENCE OF 235-665 FROM N.A.
RX MEDLINE; 93144345.
RA NAKAJIMA N.; SADO T.;
RT "Nucleotide sequence of a mouse lamin A cDNA and its deduced amino acid sequence.";
RL BIOCHIM. BIOPHYS. ACTA 1171:311-314(1993).
CC -!- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS LAYER ON THE NUCLEOLAR PLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,


```

ID PHT2_PSEPU STANDARD; PRT; 324 AA.
AC Q05192;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PHTHALATE 4,5-DIOXYGENASE OXYGENASE REDUCTASE SUBUNIT (EC 1.14.12.7).
GN PHT2.
OS PSEUDOMONAS PUTIDA.
OG PLASMID PHT.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
OC PSEUDOMONAS.
RN [1]
RC STRAIN-NH102-2.
RA NOMURA Y., NAKAGAWA M., OGAWA N., HARASHIMA S., OSHIMA Y.;
RT "Genes in PHT plasmid encoding the initial degradation pathway of
  phthalate in Pseudomonas putida.";
RL J. FERMENT. BIOENG. 74:333-344(1992).
CC -1- CATALYTIC ACTIVITY: PHTHALATE + NADH + O(2) -> CIS-4,5-
  DIHYDROXYCYCLOHEXA-1(6),2-DIENE-1,2-DICARBOXYLATE + NAD(+) +
  H(2)O.
CC -1- PATHWAY: FIRST STEP OF PHTHALATE DEGRADATION.
CC -1- COFACTOR: FMN (BY SIMILARITY).
CC -1- SUBUNIT: THIS DIOXYGENASE SYSTEM CONSISTS OF TWO PROTEINS:
  PHTHALATE OXYGENASE AND PHTHALATE OXYGENASE REDUCTASE.
CC -1- INDUCTION: INDUCED BY PHTHALATE AND REPRESSED BY GLUCOSE.
CC -1- SIMILARITY: IN THE C-TERMINAL, BELONGS TO THE 2FE2S PLANT-TYPE
  FERREDOXIN FAMILY.
CC -1- SIMILARITY: BELONGS TO THE PDR/VANB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D13229; G295709;
DR PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
DR PFAM; PF00111; fer2; 1.
DR PFAM; PF00175; oxidored_fad; 1.
DR PLASMID; AROMATIC HYDROCARBONS CATABOLISM; IRON-SULFUR; NAD;
  OXIDOREDUCTASE; ELECTRON TRANSPORT; MONOOXYGENASE; FMN; FLAVOPROTEIN.
FT NP_BIND 1 105 FMN (BY SIMILARITY).
FT NP_BIND 115 229 NAD (BY SIMILARITY).
FT METAL 275 275 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 280 280 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 283 283 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 311 311 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
SQ SEQUENCE 324 AA; 35999 MW; 216AA92F CRC32;

Query Match 76.8%; Score 43; DB 1; Length 324;
Best Local Similarity 75.0%; Pred. No. 1.94e+01;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 43 TAGAHLRV 50
:| | | | |
QY 57 SAGLHLRV 64

RESULT 11
ID RP55_RHIME STANDARD; PRT; 513 AA.
AC P33985;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE RNA POLYMERASE SIGMA-54 FACTOR (VERSION 2).
GN RPON OR NTRA.
OS RHIZOBIUM MELILOTI.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; SINORHIZOBIUM.
RN [1]

SEQUENCE FROM N.A.
RC STRAIN-104A14;
RX MEDLINE; 89359151.
RA SHATTERS R.G., SOMERVILLE J.E., KAHN M.L.;
RT "Regulation of glutamine synthetase II activity in Rhizobium meliloti
  104A14.";
RL J. BACTERIOL. 171:5087-5094(1989).
CC -1- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
  ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
  THEN IS RELEASED.
CC -1- FUNCTION: THIS SIGMA FACTOR IS RESPONSIBLE FOR THE EXPRESSION OF
  THE NITROGEN FIXATION PROTEINS (NIF OPERON), GLNA AND DCTA FOR
  DICARBOXYLATE TRANSPORT. THE OPEN COMPLEX (SIGMA-54 AND CORE RNA
  POLYMERASE) SERVES AS THE RECEPTOR FOR RECEIPT OF THE MELTING
  SIGNAL FROM THE REMOTELY BOUND ACTIVATOR PROTEINS NIFA, NTRC, OR
  DCTD FOR THE EXPRESSION OF THE REGULATED PROTEINS.
CC -1- SIMILARITY: BELONGS TO THE SIGMA-54 FACTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M28846; G152147;
DR PIR; A33966;
DR PROSITE; PS00717; SIGMA54_1; 1.
DR PROSITE; PS00718; SIGMA54_2; 1.
DR PROSITE; PS00719; SIGMA54_3; 1.
DR PFAM; PF00309; Sigma54_factors; 1.
DR KW TRANSCRIPTION REGULATION; SIGMA FACTOR; DNA-DIRECTED RNA POLYMERASE;
  DNA-BINDING; NITROGEN FIXATION.
FT DOMAIN 11 29 GLN-RICH
FT DOMAIN 21 42 LEUCINE-ZIPPER (POTENTIAL).
FT DOMAIN 190 211 LEUCINE-ZIPPER (POTENTIAL).
FT DOMAIN 394 413 H-T-H MOTIF (POTENTIAL).
FT SITE 483 491 RPON BOX.
SQ SEQUENCE 513 AA; 56821 MW; F3DE59C4 CRC32;

Query Match 76.8%; Score 43; DB 1; Length 513;
Best Local Similarity 85.7%; Pred. No. 1.94e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 4 SASLHLR 10
:| | | | |
QY 57 SAGLHLR 63

RESULT 12
ID RP54_RHIME STANDARD; PRT; 523 AA.
AC P17263;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE RNA POLYMERASE SIGMA-54 FACTOR (VERSION 1).
GN RPON OR NTRA.
OS RHIZOBIUM MELILOTI.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; SINORHIZOBIUM.
RN [1]
RC STRAIN-NH102-2.
RX MEDLINE; 87222158.
RA RONSON C.W., NIXON B.T., ALBRIGHT L.M., AUSUBEL F.M.;
RT "Rhizobium meliloti ntra (rpon) gene is required for diverse
  metabolic functions.";
RL J. BACTERIOL. 169:2424-2431(1987).
CC -1- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
  ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
  THEN IS RELEASED.
CC -1- FUNCTION: THIS SIGMA FACTOR IS RESPONSIBLE FOR THE EXPRESSION OF
  THE NITROGEN FIXATION PROTEINS (NIF OPERON), GLNA AND DCTA FOR

```

CC DICARBOXYLATE TRANSPORT. THE OPEN COMPLEX (SIGMA-54 AND CORE RNA
CC POLYMERASE) SERVES AS THE RECEPTOR FOR RECEIPT OF THE MELTING
CC SIGNAL FROM THE REMOTELY BOUND ACTIVATOR PROTEINS NIFA, NTRC, OR
CC DCTD FOR THE EXPRESSION OF THE REGULATED PROTEINS.
CC -1- SIMILARITY: BELONGS TO THE SIGMA-54 FACTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M16513; G152390; -
CC PROSITE: PS00717; SIGMA54_1; 1.
CC PROSITE: PS00718; SIGMA54_2; 1.
CC PROSITE: PS00044; SIGMA54_3; 1.
CC PFAM: PF00309; Sigma54_factors; 1.
CC KW TRANSCRIPTION REGULATION; SIGMA FACTOR; DNA-DIRECTED RNA POLYMERASE;
CC DNA-BINDING; NITROGEN FIXATION.
CC FT DOMAIN 11 29 GLN-RICH
CC FT DOMAIN 21 42 LEUCINE-ZIPPER (POTENTIAL).
CC FT DOMAIN 189 210 LEUCINE-ZIPPER (POTENTIAL).
CC FT DNA_BIND 393 412 H-T-H MOTIF (POTENTIAL).
CC FT SITE 482 490 RPN BOX.
CC SQ SEQUENCE 523 AA; 57815 MW; 2EB5FCD7 CRC32;
CC -----
CC Query Match 76.8%; Score 43; DB 1; Length 523;
CC Best Local Similarity 85.7%; Pred. No. 1.94e+01;
CC Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC Db 4 SASLHLR 10
CC |::|||
CC QY 57 SAGLHLR 63
CC -----
CC RESULT 13
CC ID RP54_RHISN STANDARD; PRT; 525 AA.
CC AC P22881;
CC DT 01-AUG-1991 (REL. 19, CREATED)
CC DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
CC DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
CC DE RNA POLYMERASE SIGMA-54 FACTOR.
CC GN RPN OR NTRC.
CC OS RHIZOBIUM SP. (STRAIN NGR234).
CC OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
CC OC RHIZOBIACEAE; RHIZOBIUM.
CC [1]
CC RN SEQUENCE FROM N.A.
CC RP MEDLINE; 91008923.
CC RX VAN SLOOTEN J.C., CERVANTES E., BROUGHTON W.J., WONG C.H., STANLEY J.;
CC "Sequence and analysis of the rpn sigma factor gene of rhizobium sp.
CC strain NGR234, a primary coregulator of symbiosis.";
CC RL J. BACTERIOL. 172:5563-5574(1990).
CC CC -1- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
CC THEN IS RELEASED.
CC CC -1- FUNCTION: THIS SIGMA FACTOR IS RESPONSIBLE FOR THE EXPRESSION OF
CC THE NITROGEN FIXATION PROTEINS (NIF OPERON), GLNA, AND DCTA FOR
CC DICARBOXYLATE TRANSPORT. THE OPEN COMPLEX (SIGMA-54 AND CORE RNA
CC POLYMERASE) SERVES AS THE RECEPTOR FOR RECEIPT OF THE MELTING
CC SIGNAL FROM THE REMOTELY BOUND ACTIVATOR PROTEINS NIFA, NTRC, OR
CC DCTD FOR THE EXPRESSION OF THE REGULATED PROTEINS.
CC CC -1- SIMILARITY: BELONGS TO THE SIGMA-54 FACTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL: M58481; G152432; -
CC PIR: A36130; A36130.
CC PROSITE: PS00717; SIGMA54_1; 1.
CC PROSITE: PS00718; SIGMA54_2; 1.
CC PROSITE: PS00044; SIGMA54_3; 1.
CC PFAM: PF00309; Sigma54_factors; 1.
CC KW TRANSCRIPTION REGULATION; SIGMA FACTOR; DNA-DIRECTED RNA POLYMERASE;
CC DNA-BINDING.
CC FT DOMAIN 11 29 GLN-RICH
CC FT DOMAIN 21 42 LEUCINE-ZIPPER (POTENTIAL).
CC FT DOMAIN 191 212 LEUCINE-ZIPPER (POTENTIAL).
CC FT DNA_BIND 393 412 H-T-H MOTIF (POTENTIAL).
CC FT SITE 484 492 RPN BOX.
CC SQ SEQUENCE 525 AA; 57800 MW; 6691BDA1 CRC32;
CC -----
CC Query Match 76.8%; Score 43; DB 1; Length 525;
CC Best Local Similarity 85.7%; Pred. No. 1.94e+01;
CC Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC Db 4 SASLHLR 10
CC |::|||
CC QY 57 SAGLHLR 63
CC -----
CC RESULT 14
CC ID DSG2_HUMAN STANDARD; PRT; 1117 AA.
CC AC Q14126;
CC DT 01-NOV-1997 (REL. 35, CREATED)
CC DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
CC DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
CC DE DESMOGLEIN 2 PRECURSOR (HDGC).
CC GN DSG2
CC OS HOMO SAPIENS (HUMAN).
CC OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
CC OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
CC [1]
CC RN SEQUENCE FROM N.A.
CC RC TISSUE-COLON CARCINOMA;
CC RX MEDLINE; 94192736.
CC RA SCHAEFER S., KOCH P.J., FRANK W.W.;
CC "Identification of the ubiquitous human desmoglein, Dsg2, and the
CC expression catalogue of the desmoglein subfamily of desmosomal
CC cadherins.";
CC RL EXP. CELL RES. 211:391-399(1994).
CC [2]
CC RN SEQUENCE OF 777-1117 FROM N.A.
CC RP MEDLINE; 92037656.
CC RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R., FRANK W.W.;
CC "Complete amino acid sequence of the epidermal desmoglein precursor
CC polypeptide and identification of a second type of desmoglein gene.";
CC RL EUR. J. CELL BIOL. 55:200-208(1991).
CC CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC CC -1- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.
CC CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. BELONGS TO THE
CC DESMOsome SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z26317; G416178; -
CC MIM: 125671; -
CC PROSITE: PS00232; CADHERIN; 3.
CC -----

DR PFAM; PF00028; cadherin; 4.
DR HSP; P15116; INCI.
KW CELL ADHESION; GLYCOPROTEIN; TRANSMEMBRANE; REPEAT; SIGNAL;
KW CYTOSKELETON; CALCIUM-BINDING.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 48 POTENTIAL.
FT CHAIN 49 1117 DESMOGLEIN 2.
FT DOMAIN 49 608 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 609 633 POTENTIAL.
FT DOMAIN 634 1117 CYTOPLASMIC (POTENTIAL).
FT REPEAT 49 159 CADHERIN 1.
FT REPEAT 160 272 CADHERIN 2.
FT REPEAT 273 387 CADHERIN 3.
FT REPEAT 388 502 CADHERIN 4.
FT REPEAT 880 911 DESMOGLEIN REPEAT 1.
FT REPEAT 912 941 DESMOGLEIN REPEAT 2.
FT REPEAT 942 967 DESMOGLEIN REPEAT 3.
FT REPEAT 968 991 DESMOGLEIN REPEAT 4.
FT REPEAT 992 1020 DESMOGLEIN REPEAT 5.
FT REPEAT 1021 1050 DESMOGLEIN REPEAT 6.
FT CARBOHYD 111 111 POTENTIAL.
FT CARBOHYD 181 181 POTENTIAL.
FT CARBOHYD 308 308 POTENTIAL.
FT CARBOHYD 461 461 POTENTIAL.
FT CARBOHYD 513 513 POTENTIAL.
SQ SEQUENCE 1117 AA; 122385 MW; 84D3B898 CRC32;

Query Match 76.8%; Score 43; DB 1; Length 1117;
Best Local Similarity 62.5%; Pred. No. 1.94e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 20 GSGHLQV 27
:||||:
Qy 57 SAGLHLRV 64

RESULT 15
ID VIT6_OSCBR STANDARD; PRT; 1660 AA.
AC Q94637;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE VITELLOGENIN 6 PRECURSOR.
GN VIT-6.
OS OSCEIUS BREVESOPHAGA.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; RHABDITINAE; OSCEIUS.
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN-CEWI;
RX MEDLINE; 96212989.
RA WINTER C.E., PENHA C., BLUMENTHAL T.;
RT "Comparison of a vitellogenin gene between two distantly related
rhabditid nematode species".
RL MOL. BIOL. EVOL. 13:674-684(1996).
CC -!- FUNCTION: PRECURSOR OF THE EGG-YOLK PROTEINS THAT ARE SOURCES OF
NUTRIENTS DURING EMBRYONIC DEVELOPMENT (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U35449; G1515337; -
DR HSP; P11064; IPR.
KW STORAGE PROTEIN; MULTIGENE FAMILY; SIGNAL.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 1660 VITELLOGENIN 6.
FT CARBOHYD 237 237 POTENTIAL.
FT CARBOHYD 383 383 POTENTIAL.

FT CARBOHYD 695 695 POTENTIAL.
FT CARBOHYD 1307 1307 POTENTIAL.
FT CARBOHYD 1596 1596 POTENTIAL.
FT CARBOHYD 1629 1629 POTENTIAL.
SQ SEQUENCE 1660 AA; 192108 MW; 9AA62413 CRC32;

Query Match 76.8%; Score 43; DB 1; Length 1660;
Best Local Similarity 71.4%; Pred. No. 1.94e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 887 TGLRLRV 893
:||||:
Qy 58 AGLHLRV 64

Search completed: Wed Sep 1 16:26:45 1999
Job time : 8 secs.

W P S R L H (TW)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:27:02 1999; MasPar time 5.10 Seconds
Tabular output not generated. 85.590 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (57-64) from PCTUS9913024.pep (10 of 12)
Perfect Score: 56
Sequence: 1 SAGLHLRV 8

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phase 10:sp-plant 11:sp-rodent 12:sp_unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 21.702; Variance 24.925; scale 0.871

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	ID	Description	Pred. No.
1	50	89.3	4293 11	O08852 POLYCYSTIC KIDNEY DISE	1.66e+00
2	48	85.7	415 10	O80705 F8K4.20 PROTEIN.	4.68e+00
3	48	85.7	479 5	O17697 SIMILAR TO GLUTAMATE R	4.68e+00
4	48	85.7	2605 2	Q50858 SAFRAMYCIN MX1 SYNTHET	4.68e+00
5	47	83.9	191 2	P97168 INTEGRASE (FRAGMENT).	7.77e+00
6	47	83.9	284 2	O05288 INTEGRASE (FRAGMENT).	7.77e+00
7	47	83.9	303 2	Q49186 INTEGRASE.	7.77e+00
8	47	83.9	336 2	Q51431 INTEGRASE (FRAGMENT).	7.77e+00
9	47	83.9	337 2	Q56438 INTEGRASE.	7.77e+00
10	47	83.9	337 2	Q56440 HYPOTHETICAL 38.4 KD P	7.77e+00
11	47	83.9	337 2	Q48337 INTEGRASE.	7.77e+00
12	47	83.9	337 2	O50324 INTEGRASE.	7.77e+00
13	45	80.4	238 14	P90337 REPLICATION-RELATED PR	2.10e+01
14	45	80.4	260 2	O68996 HYPOTHETICAL 27.8 KD P	2.10e+01
15	45	80.4	499 2	O07561 HYPOTHETICAL 54.4 KD P	2.10e+01
16	45	80.4	755 14	P89107 SCVP86.	2.10e+01
17	45	80.4	2133 14	O98203 MC035R.	2.10e+01
18	44	78.6	438 8	O80030 RIBULOSE BISPHOSPHATE	3.41e+01
19	44	78.6	440 8	Q32141 RIBULOSE BISPHOSPHATE	3.41e+01
20	44	78.6	440 8	O78589 RIBULOSE BISPHOSPHATE	3.41e+01

21	44	78.6	447	8	O78577	RIBULOSE BISPHOSPHATE	3.41e+01
22	44	78.6	447	8	Q37324	RIBULOSE BISPHOSPHATE	3.41e+01
23	44	78.6	447	8	O96083	RIBULOSE BISPHOSPHATE	3.41e+01
24	44	78.6	466	8	P93933	RIBULOSE BISPHOSPHATE	3.41e+01
25	44	78.6	467	8	Q37330	RIBULOSE BISPHOSPHATE	3.41e+01
26	43	76.8	341	1	O29051	CONSERVED HYPOTHETICAL	5.49e+01
27	43	76.8	490	5	Q19324	CODED FOR BY C. ELEGAN	5.49e+01
28	43	76.8	496	5	Q27693	CYTCHROME P450 (EC 1.	5.49e+01
29	43	76.8	514	2	O59751	RNA POLYMERASE SIGMA-5	5.49e+01
30	43	75.8	884	10	O04085	UNKNOWN PROTEIN.	5.49e+01
31	42	75.0	170	2	O52584	HYPOTHETICAL 17.7 KD P	8.78e+01
32	42	75.0	216	8	O48269	CYTCHROME C OXIDASE S	8.78e+01
33	42	75.0	321	4	O99608	NECDIN RELATED PROTEIN	8.78e+01
34	42	75.0	409	14	Q64961	NUCLEOCAPSID PROTEIN.	8.78e+01
35	42	75.0	442	13	Q73803	FUGU HEDGEHOG.	8.78e+01
36	42	75.0	453	4	O14586	BAC CLONE GS303P24 FRO	8.78e+01
37	42	75.0	494	10	Q39520	LEGUMIN (FRAGMENT).	8.78e+01
38	42	75.0	502	10	O40689	LEGUMIN.	8.78e+01
39	42	75.0	561	3	O60051	COHESIN.	8.78e+01
40	42	75.0	689	13	O13265	TISSUE TRANSGLUTAMINAS	8.78e+01
41	42	75.0	713	3	O42947	HYPOTHETICAL 80.1 KD P	8.78e+01
42	42	75.0	849	5	O77068	ALPHA ACTININ (FRAGMEN	8.78e+01
43	42	75.0	1072	6	O77690	OLFACTORY ENTEROTOXIN	8.78e+01
44	42	75.0	1262	11	Q60988	TALI INTERRUPTING LOCUS	8.78e+01
45	42	75.0	1706	11	Q63755	RETINOBLASTOMA PROTEIN	8.78e+01

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	4293 AA.
ID	O08852			
AC	O08852:			
DT	01-JUL-1997 (TREMBLREL. 04, CREATED)			
DT	01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	POLYCYSTIC KIDNEY DISEASE 1 PROTEIN.			
GN	PKD1.			
OS	MUS MUSCULUS (MOUSE).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;			
OC	SCUROGNATHI; MURIDAE; MURINAE; MUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 97262094.			
RA	LOHNING C., NOWICKA U., FRISCHAUF A.M.;			
RT	*The mouse homolog of PKD1: sequence analysis and alternative			
RT	splicing.*;			
RL	MAMM. GENOME 8:307-311(1997).			
DR	EMBL; U70209; G2138183; -			
DR	PFAM; PF00560; LRR; 1.			
DR	PFAM; PF00801; PKD; 16.			
SQ	SEQUENCE 4293 AA; 466545 MW; 10E37A8A CRC32;			

Query Match 89.3%; Score 50; DB 11; Length 4293;
Best Local Similarity 85.7%; Pred. No. 1.66e+00;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 2910 AGLHLRI 2916
|||||
Qy 58 AGLHLRV 64

RESULT	2	PRELIMINARY:	PRT:	415 AA.
ID	O80705			
AC	O80705:			
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	F8K4.20 PROTEIN.			
GN	F8K4.20.			
OS	ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).			
OC	EUKARYOTA; VIRIDIPANTAE; EMBRYOPHYTA; TRACHEOPHYTA;			
OC	EUPHYLLIOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; ROSIDAE;			
OC	CAPPARALES; BRASSICACEAE; ARABIDOPSIS.			

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA VYOTSKAIA V.S., SCHWARTZ J.R., TORIUMI M., KWAN A., YU G., OJI O.,
LIU S., LI J., ARAUJO R., AU M., BRENDEL V., BUEHLER E., CONWAY A.B.,
CONWAY A.R., DEWAR K., FENG J., KIM C., KURTZ D., LI Y., PALM C.J.,
SHANN P., SUN H., DAVIS R.W., ECKER J.R., FEDERSPIEL N.A.,
RA THEOLOGIS A.;
RT "Arabidopsis thaliana chromosome 1 BAC F8K4 sequence.";
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA THEOLOGIS A.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA THEOLOGIS A.;
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA THEOLOGIS;
RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AC004392; G3367531;
SQ SEQUENCE 415 AA; 46438 MW; B5E0FFEL CRC32;
Query Match 85.7%; Score 48; DB 10; Length 415;
Best Local Similarity 100.0%; Pred. No. 4.68e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 14 GLHLRV 19
QY 59 GLHLRV 64
RESULT 3
ID Q17697 PRELIMINARY; PRT; 479 AA.
AC Q17697;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE SIMILAR TO GLUTAMATE RECEPTORS.
GN COB8.9.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
BOFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA LEIMBACH D.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
```

```
RN [3]
RP SEQUENCE FROM N.A.
RL WATERSTON R.;
RA SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U39849; G1055049;
DR PFAM; PF00060; 11g_chan; 2.
SQ SEQUENCE 479 AA; 54976 MW; 604C46C9 CRC32;
Query Match 85.7%; Score 48; DB 5; Length 479;
Best Local Similarity 100.0%; Pred. No. 4.68e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 46 GLHLRV 51
QY 59 GLHLRV 64
RESULT 4
ID Q50858 PRELIMINARY; PRT; 2605 AA.
AC Q50858;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE SAFRAMYCIN MX1 SYNTHETASE A.
GN SAFA.
OS MYXOCOCCUS XANTHUS.
OC BACTERIA; PROTEOBACTERIA; DELTA SUBDIVISION; MYXOBACTERIA;
OC MYXOCOCCALES; CYSTOBACTERINEAE; MYXOCOCCACEAE; MYXOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DM504-15;
RX MEDLINE; 96032006.
RA POSPIECH A., CLUZEL B., BIETENHADER J., SCHUPP T.;
RT "A new Myxococcus xanthus gene cluster for the biosynthesis of the
antibiotic saframycin Mx1 encoding a peptide synthetase.";
RL MICROBIOLOGY 141:1793-1803(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-DM504-15;
RX MEDLINE; 97090395.
RA POSPIECH A., BIETENHADER J., SCHUPP T.;
RT "two multifunctional peptide synthetases and an O-methyltransferase
are involved in the biosynthesis of the DNA-binding antibiotic and
antitumour agent saframycin Mx1 from Myxococcus xanthus.";
RL MICROBIOLOGY 142:741-746(1996).
DR EMBL; U24657; G1171129;
DR PROSITE; PS00455; AMP_BINDING; 2.
DR PFAM; PF00501; AMP-binding; 2.
DR PFAM; PF00550; pp-binding; 2.
DR PFAM; PF00668; DUF4; 2.
SQ SEQUENCE 2605 AA; 285745 MW; 09C34B84 CRC32;
Query Match 85.7%; Score 48; DB 2; Length 2605;
Best Local Similarity 100.0%; Pred. No. 4.68e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1530 GLHLRV 1535
QY 59 GLHLRV 64
RESULT 5
ID P97168 PRELIMINARY; PRT; 191 AA.
AC P97168;
DT 01-MAY-1997 (TREMREL. 03, CREATED)
DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE INTEGRASE (FRAGMENT).
GN IN1.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
```

RP SEQUENCE FROM N.A.

RC STRAIN-RA33.2;
RA ADRIAN P.V., THOMSON C.J., KLUGMAN K.P., AMYES S.G.B.;
RL SUBMITTED (AUG-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: 250802; G950043;
FT NON_TER 191
SQ SEQUENCE 191 AA; 21891 MW; C007CICE CRC32;

Query Match 83.9%; Score 47; DB 2; Length 191;
Best Local Similarity 75.0%; Pred. No. 7.77e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLQRLV 155
QY 57 SAGLHLRV 64

RESULT 6

ID O05288 PRELIMINARY; PRT; 284 AA.
AC O05288;
DT 01-JUL-1997 (TREMREL. 04, CREATED)
DT 01-JUL-1997 (TREMREL. 04, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE INTEGRASE (FRAGMENT).
OS KLEBSIELLA OXYTOCA.
OG PLASMID PACM1.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC KLEBSIELLA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97312761.
RA PRESTON K.E.; KACICA M.A., LIMBERGER R.J., ARCHINAL W.A.,
RA VENEZIA R.A.;
RT "The resistance and integrase genes of pACM1, a conjugative
multiple-resistance plasmid, from Klebsiella oxytoca."
RL PLASMID 37:105-118(1997).
DR EMBL: U09445; G1906652;
DR PFAM: PF00589; Phage_integrase; 1.
KW PLASMID.
FT NON_TER 284
SQ SEQUENCE 284 AA; 32632 MW; 6DSALB7F CRC32;

Query Match 83.9%; Score 47; DB 2; Length 284;

Best Local Similarity 75.0%; Pred. No. 7.77e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLQRLV 155
QY 57 SAGLHLRV 64

RESULT 7

ID Q49186 PRELIMINARY; PRT; 303 AA.
AC Q49186;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE INTEGRASE.
OS MYCOBACTERIUM FORTUITUM.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; CORYNEBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX TRANSPOSOM-TN10;
RX MEDLINE; 90294910.
RA MARTIN C., TIMM J., RAUZIER J., GOMEZ-LUS R., DAVIES J., GICQUEL B.;
RT "Transposition of an antibiotic resistance element in mycobacteria."
RL NATURE 345:739-743(1990).
DR EMBL: X53635; G4286;
DR PFAM: PF00589; Phage_integrase; 1.
SQ SEQUENCE 303 AA; 34884 MW; iC45D6BB CRC32;

Query Match 83.9%; Score 47; DB 2; Length 303;

Best Local Similarity 75.0%; Pred. No. 7.77e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLQRLV 155
QY 57 SAGLHLRV 64

RESULT 8

ID Q51431 PRELIMINARY; PRT; 336 AA.
AC Q51431;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE INTEGRASE (FRAGMENT).
GN TNPI.
OS PSEUDOMONAS AERUGINOSA.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
OC PSEUDOMONAS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-101;
RA LARAKI N., ROSSOLINI G., GALLEN M., FRERE J.M.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: X98393; E248569;
DR PFAM: PF00589; Phage_integrase; 1.
FT NON_TER 336
SQ SEQUENCE 336 AA; 38239 MW; 420D38E3 CRC32;

Query Match 83.9%; Score 47; DB 2; Length 336;

Best Local Similarity 75.0%; Pred. No. 7.77e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLQRLV 155
QY 57 SAGLHLRV 64

RESULT 9

ID Q56438 PRELIMINARY; PRT; 337 AA.
AC Q56438;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE INTEGRASE.
GN TNPI.
OS SHIGELLA SONNEI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SHIGELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSOM-TN2426;
RX MEDLINE; 86041899.
RA FLING M.E., KOFF J., RICHARDS C.;
RT "Nucleotide sequence of the transposon Tn7 gene encoding an
aminoglycoside-modifying enzyme, 3'(9)-O-nucleotidyltransferase."
RL NUCLEIC ACIDS RES. 13:7095-7106(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC TRANSPOSOM-TN2426;
RA WIEDEMANN B.;
RL ANTIMICROB. AGENTS CHEMOTHER. 0:85-92(1986).
RN [3]
RP SEQUENCE FROM N.A.
RC TRANSPOSOM-TN2426;
RX MEDLINE; 87066734.
RA CAMERON F.H., GROOT OBBINK D.J., ACKERMAN V.P., HALL R.M.;
RT "Nucleotide sequence of the AAD(2'') aminoglycoside
adenylyltransferase determinant aadB. Evolutionary relationship of
this region with those surrounding aadA in R538-1 and dhfrII in
R389."
RL NUCLEIC ACIDS RES. 14:8625-8635(1986).
RN [4]


```

RP SEQUENCE FROM N.A.
RC TRANSPOSON-TN2426;
RX MEDLINE: 88086917.
RA TENOVER F.C., FILPULA D., PHILLIPS K.L., FLORDE J.J.;
RT "Cloning and sequencing of a gene encoding an aminoglycoside
RL 6'-N-acetyltransferase from an R factor of Citrobacter diversus.";
RN J. BACTERIOL. 170:471-473(1988).
RN [5]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-TN2426;
RX MEDLINE: 90299796.
RA MERCIER J., LACHAPPELLE J., COUTURE F., LAFOND M., VEZINA G.,
BOISSINOT M., LEVESQUE R.C.;
RT "Structural and functional characterization of tnpI, a recombinase
RL locus in Tn21 and related beta-lactamase transposons.";
RN J. BACTERIOL. 172:3745-3757(1990).
RN [6]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-TN2426;
RX MEDLINE: 93329367.
RA ZUHLSDORF M.T., WIEDEMANN B.;
RT "Functional and physiological characterization of the Tn21 cassette
RL for resistance genes in Tn2426.";
RN J. GEN. MICROBIOL. 139:995-1002(1993).
DR EMBL: M86913; GI55024; -.
DR PFAM: PF00589; Phage_integrase; 1.
SQ SEQUENCE 337 AA; 38415 MW; 093D245D CRC32;

Query Match 83.9%; Score 47; DB 2; Length 337;
Best Local Similarity 75.0%; Pred. No. 7.77e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155
QY 57 SAGLHLRV 64

RESULT 10
ID Q56440 PRELIMINARY; PRT; 337 AA.
AC Q56440;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 38.4 KD PROTEIN.
OS UNKNOWN.
OC BACTERIA; ENVIRONMENTAL SAMPLES.
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-TN21;
RX MEDLINE: 90299796.
RA MERCIER J., LACHAPPELLE J., COUTURE F., LAFOND M., VEZINA G.,
BOISSINOT M., LEVESQUE R.C.;
RT "Structural and functional characterization of tnpI, a recombinase
RL locus in Tn21 and related beta-lactamase transposons.";
RN J. BACTERIOL. 172:3745-3757(1990).
DR EMBL: M33633; GI197009; -.
DR PFAM: PF00589; Phage_integrase; 1.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 337 AA; 38376 MW; D8B33A67 CRC32;

Query Match 83.9%; Score 47; DB 2; Length 337;
Best Local Similarity 75.0%; Pred. No. 7.77e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155
QY 57 SAGLHLRV 64

RESULT 11
ID Q48337 PRELIMINARY; PRT; 337 AA.
AC Q48337; P71194; Q48386;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

```

```

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE INTEGRASE.
GN INT.
OS ESCHERICHIA COLI, AND KLEBSIELLA AEROGENES.
OG PLASMID PDG0100, AND PLASMID INCP-BETA R751.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87066734.
RA CAMERON F.H., GROOT OBBINK D.J., ACKERMAN V.P., HALL R.M.;
RT "Nucleotide sequence of the AAD(2'') aminoglycoside
adenylyltransferase determinant aadB. Evolutionary relationship of
this region with those surrounding aadA in R538-I and dhfrII in
R388.";
RL NUCLEIC ACIDS RES. 14:8625-8635(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88015610.
RA HALL R.M., VOCKLER C.;
RT "The region of the IncN plasmid R46 coding for resistance to
beta-lactam antibiotics, streptomycin/spectinomycin and sulphonamides
is closely related to antibiotic resistance segments found in IncW
plasmids and in Tn21-like transposons.";
RL NUCLEIC ACIDS RES. 15:7491-7501(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90272804.
RA HALL R.M., STOKES H.W.;
RT "The structure of a partial duplication in the integron of plasmid
PDG0100.";
RL PLASMID 23:76-79(1990).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92206862.
RA PARSONS Y., HALL R.M., STOKES H.W.;
RT "A new trimethoprim resistance gene, dhfrX, in the In7 integron of
plasmid PDG0100.";
RL ANTIMICROB. AGENTS CHEMOTHER. 35:2436-2439(1991).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93391548.
RA STOKES H.W., TOMARAS C., PARSONS Y., HALL R.M.;
RT "The partial 3'-conserved segment duplications in the integrons In6
from pSa and In7 from PDG0100 have a common origin.";
RL PLASMID 30:39-50(1993).
RN [6]
RP SEQUENCE FROM N.A.
RC SPECIES-AEROBACTER AEROGENES;
RX MEDLINE: 85215496.
RA SMITH C.A., THOMAS C.M.;
RT "Comparison of the nucleotide sequences of the vegetative replication
origins of broad host range Incp plasmids R751 and RK2 reveals
conserved features of probable functional importance.";
RL NUCLEIC ACIDS RES. 13:557-572(1985).
RN [7]
RP SEQUENCE FROM N.A.
RA SUNDBLOM L.;
RT SUBMITTED (MAR-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [8]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94252994.
RA RADSTROM P., SKOLD O., SWEDBERG G., FLENSBURG J., ROY P.H.,
SUNDBLOM L.;
RT "Transposon Tn5090 of plasmid R751, which carries an integron, is
related to Tn7, Mu, and the retroelements.";
RL J. BACTERIOL. 176:3257-3268(1994).
DR EMBL: L06418; GI49117; -.
DR EMBL: U67194; GI572565; -.
DR EMBL: X72585; G288634; -.
DR PFAM: PF00589; Phage_integrase; 1.
KW PLASMID.

```

SQ SEQUENCE 337 AA; 38358 MW; D25371F2 CRC32;

Query Match 83.9%; Score 47; DB 2; Length 337;

Best Local Similarity 75.0%; Pred. No. 7.77e+00;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155

QY 57 SAGLHLRV 64

RESULT 12
ID O50324 PRELIMINARY; PRT; 337 AA.

AC O50324;

DT 01-JUN-1998 (TREMBLREL. 06, CREATED)

DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE INTEGRASE.

GN INT.

OS PSEUDOMONAS AERUGINOSA.

OG PLASMID PMS350.

OC BACTERIA; PROTEOBACTERIA: GAMMA SUBDIVISION; PSEUDOMONAS GROUP;

OC PSEUDOMONAS.

RN [1]

RP SEQUENCE FROM N.A.

RX PLASMID-PMS350;

RA IYOBE S., MINAMI S., YAMADA H.;

RT "Insertion of a carboxypeptidase gene cassette into an integron of a

RT Pseudomonas aeruginosa plasmid[published erratum appears in J

RT Antimicrob Chemother 1997 Jun;39(6):845].";

RL J. ANTIMICROB. CHEMOTHER. 38:1114-1115(1996).

DR EMBL; D78374; D1024672; -

SQ SEQUENCE 337 AA; 38298 MW; 753AD743 CRC32;

Query Match

Best Local Similarity 83.9%; Score 47; DB 2; Length 337;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 148 SEGQLRV 155

QY 57 SAGLHLRV 64

RESULT 13
ID P90337 PRELIMINARY; PRT; 238 AA.

AC P90337;

DT 01-MAY-1997 (TREMBLREL. 03, CREATED)

DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)

DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE REPLICATION-RELATED PROTEIN.

GN SCVP26.

OS SAGUARO CACTUS VIRUS.

OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; TOMBUSVIRIDAE;

OC CARMOVIRUS.

RN [1]

RP SEQUENCE FROM N.A.

RA XIONG Z., WENG Z.;

RL J. GEN. VIROL. 0:0-0(0).

RN [2]

RP SEQUENCE FROM N.A.

RA XIONG Z., WENG Z.;

RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; U72332; G1685120; -

SQ SEQUENCE 238 AA; 26481 MW; 446D4179 CRC32;

Query Match

Best Local Similarity 80.4%; Score 45; DB 14; Length 238;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 23 AGLQRI 29

QY 58 AGLHLRV 64

RESULT 14

ID O68996 PRELIMINARY; PRT; 260 AA.

AC O68996;

DT 01-AUG-1998 (TREMBLREL. 07, CREATED)

DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)

DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE HYPOTHETICAL 27.8 KD PROTEIN.

OS MYCOBACTERIUM AVIUM.

OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;

OC ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-2151;

RA ECKSTEIN T.M., SILBAO F.S., INAMINE I.M., BELISLE J.T.;

RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; AF060182; G3170647; -

KW HYPOTHETICAL PROTEIN.

SQ SEQUENCE 260 AA; 27809 MW; 882DB702 CRC32;

Query Match

Best Local Similarity 80.4%; Score 45; DB 2; Length 260;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 213 SSGLYLRV 220

QY 57 SAGLHLRV 64

RESULT 15

ID O07561 PRELIMINARY; PRT; 499 AA.

AC O07561;

DT 01-JUL-1997 (TREMBLREL. 04, CREATED)

DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE HYPOTHETICAL 54.4 KD PROTEIN.

GN YHJG.

OS BACILLUS SUBTILIS.

OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;

OC BACILLUS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-168;

RA NOBACK M.A., TERPSTRA P., HOLSAPPEL S., VENEMA G., BRON S.;

RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-168;

RX MEDLINE; 98044033.

RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.;

RA AZEVEDO V., BERTERO M.G., BESIETTES P., BOLOTIN A., BORCHERT S.;

RA BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.;

RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.;

RA CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.;

RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMERSON P.T.;

RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.;

RA FRITZ C., FUJITA M., FUJITA Y., FUNA S., GALIZZI A., GALLERON N.;

RA GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.;

RA GUISEPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.;

RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.;

RA JORIS B., KARAMATA D., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C.;

RA KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUNANO M.;

RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.;

RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGEC C.;

RA MEYNE D., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.;

RA NEONE D., O'REILLY M., OGAWA K., OGIWARA A., OUDEGA B., PARK S.H.;

RA PARRO V., POHL T.M., PORTETELLE D., POROLLIK S., PRESCOTT A.M.;

RA PRESCAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.;

RA RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADATE Y.;

RA SATO T., SCANLAN E., SCHLEICH S., SCHROTER R., SCOFFONE F.;

RA SERIGUCHI J., SEKOWSKA A., SERO S.J., SERROR P., SHIN B.S., SOLDI B.;

RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.;

RA TAKEUCHI M., TAKAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
 RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
 RA VIARI A., WAMBUUT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
 RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
 RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 subtilis.";
 RL NATURE 390:249-256(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: Y14081; E324980; -;
 DR EMBL: Z99109; E1183052; -;
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 499 AA; 54356 MW; 86F6D7E9 CRC32;

Query Match 80.4%; Score 45; DB 2; Length 499;
 Best Local Similarity 100.0%; Pred. No. 2.10e+01;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 352 AGLHLR 357
 QY 58 AGLHLR 63
 |||||

Search completed: Wed Sep 1 16:27:28 1999
 Job time : 26 secs.

QY 70 GALVDSKSYA 79

RESULT 2

ID W76238 standard; protein; 146 AA.

DT 02-DEC-1998 (first entry)

DE Bacterial periplasmic binding protein fragment #13.

KW Protein-ligand binding pocket; PLBP; binding protein; ligand; modulator;

KW bacterial periplasmic binding protein; interaction energies; ischaemia;

KW basis set molecules; BSM; ionotropic glutamate receptors; treatment;

KW neuroprotectant; stroke; epilepsy; neuropathic pain.

OS Prokaryota.

PN W09838208-A2.

PD 03-SEP-1998.

PF 27-FEB-1998; U03951.

PR 28-FEB-1997; US-808804.

PA (BEAR-) BEARDSDEN BIO INC.

PI Surgess M;

DR WPI; 98-495386/42.

PT Designing a model for a ligand-binding pocket in a protein and its

PT use for assessing ligand affinity - without the need to prepare test

PT ligand, for identifying selective antagonists for ionotropic

PT glutamate receptors, potentially useful as neuroprotectants with

PT reduced side effects

PS Disclosure: Page 215-216; 218pp; English.

CC W76238-W76250 are protein fragments used in the design of a model of a

CC protein-ligand binding pocket (PLBP) of a binding protein. A model is

CC constructed based on topographic similarity to the binding pocket of a

CC bacterial periplasmic binding protein. The model is refined by energy

CC minimisation, with a high affinity ligand in the binding pocket.

CC Interaction energies of basis set molecules (BSM) are obtained by

CC calculating energy of the models for the binding pocket BP and BSM

CC individually (Er and Em) and total energy (Erm) of bound complexes

CC formed. The model is used to predict the selectivity of a potential

CC ligand for a set of related PLBP's, specifically ionotropic glutamate

CC receptors. Identified ligands are potentially useful for studying

CC receptor binding and activity and as modulators of receptor activity

CC (or lead compounds for developing such compounds). The ligands are

CC potentially useful therapeutically, e.g. as neuroprotectants during

CC ischaemia and for treatment of stroke, head injuries, epilepsy,

CC neuropathic pain etc. The method allows affinity of ligands to be

CC estimated without having to prepare them and then test them, in vitro.

CC Ligands with high selectivity for particular glutamate receptors

CC should have fewer side effects than known receptor antagonists.

SQ Sequence 146 AA;

Query Match 60.5%; Score 52; DB 35; Length 146;

Best Local Similarity 60.0%; Pred. No. 3.71e+01;

Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 96 gglidskgyg 105

QY 70 GALVDSKSYA 79

RESULT 3

ID R63070 standard; Protein; 849 AA.

AC R63070;

DT 22-MAR-1995 (first entry)

DE Human EAA3d excitatory amino acid receptor.

KW excitatory amino acid receptor; human EAA3 receptor; kainate-type;

KW EAA3d receptor; central nervous system receptor; CNS.

OS Homo sapiens.

FH Key

FT Location/Qualifiers

FT peptide 1..30

FT /label= signal_peptide

FT protein 31..849

FT /label= mature_EAA3c_receptor

FT /note= "amino acids 9-65 of the mature EAA3a

FT receptor are deleted in EAA3d"

PN CA2110933-A.

PD 12-JUN-1994.

QY 70 GALVDSKSYA 79

RESULT 4

ID R63069 standard; Protein; 865 AA.

AC R63069;

DT 22-MAR-1995 (first entry)

DE Human EAA3c excitatory amino acid receptor.

KW excitatory amino acid receptor; human EAA3 receptor; kainate-type;

KW EAA3c receptor; central nervous system receptor; CNS.

OS Homo sapiens.

FH Key

FT Location/Qualifiers

FT peptide 1..30

FT /label= signal_peptide

FT protein 31..865

FT /label= mature_EAA3c_receptor

FT /note= "truncated version of EAA3a in

FT which 40 amino acids have been eliminated

FT from the C-terminus and in which the last

FT 11 amino acids differ from those in the

FT corresponding region of EAA3a"

FT domain 31..562

FT /label= N-terminal_domain

FT /note= "extracellular"

FT region 563..840

FT /label= hydrophobic_region

FT domain 563..582

FT /label= TM-1

FT /note= "transmembrane domain"

FT domain 604..624

FT /label= TM-2

FT /note= "transmembrane domain"

FT domain 635..653

FT /label= TM-3

FT /note= "transmembrane domain"

FT domain 820..840

FT /label= TM-4

FT /note= "transmembrane domain"

FT domain 841..865

FT /label= C-terminal_domain

FT /note= "extracellular"

PN CA2110933-A.

PF 08-DEC-1993; 110933.

PR 11-DEC-1992; US-989793.

PA (KAMB/) KAMBOJ R.

PA (ELLI/) ELLIOT C E.

PA (NUTT/) NUTT S L.

PI Elliot CE, Kamboj R, Nutt SL;

DR N-PSDB; Q81158.

DR N-PSDB; Q81158.

PT Polynucleotide encoding a human excitatory amino acid 3 receptor

PT or fragment - used to assay test ligands for their interaction

PT with a human CBS receptor.

PS Claim 13; Fig 1 and Fig 4C; 35pp; English.

CC The human EAA3a receptor (R60112) and its variants EAA3b, EAA3c and

CC EAA3d (R63068-R63070) occur naturally in human brain. They are all

CC members of the EAA3 receptor family, having extracellular N- and C-

CC termini and 4 internal hydrophobic domains which anchor the receptor

CC within the cell surface membrane. The receptors bind glutamate and

CC also exhibit binding properties characteristic of kainate-type EAA

CC receptors.

CC N.B. The cDNA and corresp. amino acid sequences for EAA3b, 3c and

CC 3d do not appear in full in the specification; the sequences have

CC been compiled by combining sequences in figures 1 and 4 according

CC to the description given in the disclosure.

SQ Sequence 849 AA;

Query Match 60.5%; Score 52; DB 11; Length 849;

Best Local Similarity 60.0%; Pred. No. 3.71e+01;

Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 700 gglidskgyg 709

QY 70 GALVDSKSYA 79

RESULT 4

ID R63069 standard; Protein; 865 AA.

AC R63069;

DT 22-MAR-1995 (first entry)

DE Human EAA3c excitatory amino acid receptor.

KW excitatory amino acid receptor; human EAA3 receptor; kainate-type;

KW EAA3c receptor; central nervous system receptor; CNS.

OS Homo sapiens.

FH Key

FT Location/Qualifiers

FT peptide 1..30

FT /label= signal_peptide

FT protein 31..865

FT /label= mature_EAA3c_receptor

FT /note= "truncated version of EAA3a in

FT which 40 amino acids have been eliminated

FT from the C-terminus and in which the last

FT 11 amino acids differ from those in the

FT corresponding region of EAA3a"

FT domain 31..562

FT /label= N-terminal_domain

FT /note= "extracellular"

FT region 563..840

FT /label= hydrophobic_region

FT domain 563..582

FT /label= TM-1

FT /note= "transmembrane domain"

FT domain 604..624

FT /label= TM-2

FT /note= "transmembrane domain"

FT domain 635..653

FT /label= TM-3

FT /note= "transmembrane domain"

FT domain 820..840

FT /label= TM-4

FT /note= "transmembrane domain"

FT domain 841..865

FT /label= C-terminal_domain

FT /note= "extracellular"

PN CA2110933-A.

PD 12-JUN-1994.
 PF 08-DEC-1993; 110933.
 PR 11-DEC-1992; US-989793.
 PA (KAMB/) KAMBOJ R.
 PA (ELLI/) ELLIOT C E.
 PA (NUTT/) NUTT S L.
 PI Elliot CE, Kamboj R, Nutt SL;
 DR WPI: 94-255829/32.
 DR N-PSDB: Q81157.
 PT Polynucleotide encoding a human excitatory amino acid 3 receptor
 PT or fragment - used to assay test ligands for their interaction
 PT with a human CBS receptor.
 PS Claim 13: Fig 1 and Fig 4B; 35pp; English.
 CC The human EAA3a receptor (R60112) and its variants EAA3b, EAA3c and
 CC EAA3d (R63068-R63070) occur naturally in human brain. They are all
 CC members of the EAA3 receptor family, having extracellular N- and C-
 CC termini and 4 internal hydrophobic domains which anchor the receptor
 CC within the cell surface membrane. The receptors bind glutamate and
 CC also exhibit binding properties characteristic of kainate-type EAA
 CC receptors.
 CC N.B. the cDNA and corresp. amino acid sequences for EAA3b, 3c and
 CC 3d do not appear in full in the specification; the sequences have
 CC been compiled by combining sequences in Figures 1 and 4 according
 CC to the description given in the disclosure.
 SQ Sequence 865 AA;
 Query Match 60.5%; Score 52; DB 11; Length 865;
 Best Local Similarity 60.0%; Pred. No. 3.71e+01;
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 Db 756 gglidskgyg 765
 I:::||||:
 Qy 70 GALVDSKSYA 79
 RESULT 5
 ID R11994 standard; Protein; 901 AA.
 AC R11994;
 DT 31-JUL-1991 (first entry)
 DE Glutamate receptor 6.
 KW Glutamate receptor 6; probe; ligand; drug screening.
 OS Rattus rattus.
 PN WO9105648-A.
 PD 16-MAY-1991.
 PF 25-OCT-1990; U06153.
 PR 27-OCT-1989; US-428116.
 PA (SALK) SALK INST FOR BIOL STUD.
 PI Heinemann SF, Boulter JR, Hollmann M, Bettler B, Jensen JE;
 DR WPI: 91-164197/22.
 DR N-PSDB: Q11854.
 PT Glutamate receptors - used to screen for functional ligands and
 PT identify and isolate further receptors
 PS Disclosure; Fig 11; 109pp; English.
 CC GUR6 has a Mr of 100,000.
 CC The gene and protein can be used in drug screening, to
 CC determine whether a substance is a functional ligand for the
 CC receptor by monitoring ion channel activity.
 CC See also Q11849-855.
 SQ Sequence 901 AA;
 Query Match 60.5%; Score 52; DB 2; Length 901;
 Best Local Similarity 60.0%; Pred. No. 3.71e+01;
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 Db 756 gglidskgyg 765
 I:::||||:
 Qy 70 GALVDSKSYA 79
 RESULT 6
 ID R75883 standard; Protein; 905 AA.
 AC R75883;
 DT 10-JUL-1996 (first entry)

DE Human EAA3 receptor (Q-591).
 KW Glutamate receptor; EAA3 receptor; excitatory amino acid;
 OS Homo sapiens.
 PH Key Location/Qualifiers
 FT peptide 1..30
 FT /label= Sig_peptide
 FT 31..905
 FT /label= Mat_protein
 PN WO9517508-A2.
 PD 29-JUN-1995.
 PF 21-DEC-1994; CA0705.
 PR 23-DEC-1993; US-172188.
 PA (ALLX) ALLELIX BIOPHARMACEUTICALS INC.
 PI Kamboj R, Nutt S;
 DR WPI: 95-240670/31.
 DR N-PSDB: Q91231.
 PT Identification of human CNS receptor ligand - and identification of
 PT agents that modulate editing of human CNS receptors
 PS Claim 4; Fig 5; 59pp; English.
 CC The amino acid sequence (R75883) of human EAA3 receptor deduced from
 CC a genomic DNA sequence (Q91231) differed from that deduced from
 CC EAA3 cDNA at position 591 of the mature protein; the genomic sequence
 CC coded for glutamine at this position, and the cDNA sequence for
 CC arginine. This was due to a single nucleotide change of A to G at
 CC position 2279 as a result of RNA editing. The edited and unedited
 CC receptors were functionally distinct and appeared to be selectively
 CC expressed. CNS receptors subject to such editing (see also R75882
 CC and R75884) have value in screening cpds. for potential therapeutic
 CC utility.
 SQ Sequence 905 AA;
 Query Match 60.5%; Score 52; DB 16; Length 905;
 Best Local Similarity 60.0%; Pred. No. 3.71e+01;
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 Db 756 gglidskgyg 765
 I:::||||:
 Qy 70 GALVDSKSYA 79
 RESULT 7
 ID R63068 standard; Protein; 905 AA.
 AC R63068;
 DT 22-MAR-1995 (first entry)
 DE Human EAA3b excitatory amino acid receptor.
 KW excitatory amino acid receptor; human EAA3 receptor; kainate-type;
 KW EAA3a receptor; central nervous system receptor; CNS.
 OS Homo sapiens.
 PH Key Location/Qualifiers
 FT peptide 1..30
 FT /label= signal_peptide
 FT 31..905
 FT /label= mature_EAA3b_receptor
 FT /note= "differs from EAA3a by one amino acid,
 FT i.e. Asp at position +639 of mature
 FT EAA3a is replaced by Asn in EAA3b"
 FT 31..562
 FT /label= N-terminal_domain
 FT /note= "extracellular"
 FT 563..840
 FT /label= hydrophobic_region
 FT 563..582
 FT /label= TM-1
 FT /note= "transmembrane domain"
 FT 604..624
 FT /label= TM-2
 FT /note= "transmembrane domain"
 FT 635..653
 FT /label= TM-3
 FT /note= "transmembrane domain"
 FT 820..840
 FT /label= TM-4

```

FT FT domain /note= "transmembrane domain"
FT FT 841...905
FT FT /label= C-terminal_domain
FT FT /note= "extracellular"
PN CA2110933-A.
PD 12-JUN-1994.
PF 08-DEC-1993; 110933.
PR 11-DEC-1992; US-989793.
PA (KAMB/) KAMBOJ R.
PA (ELLI/) ELLIOT C E.
PA (NUTT/) NUTT S L.
PI Elliot CE, Kamboj R, Nutt SL;
DR WPI: 94-255829/32.
DR N-PSDB: Q70097.
PT Polynucleotide encoding a human excitatory amino acid 3 receptor
PT or fragment - used to assay test ligands for their interaction
PT with a human CBS receptor.
PS Claim 13; Fig 1 and Fig 4A; 35pp; English.
CC The human EAA3a receptor (R60112) and its variants EAA3b, EAA3c and
CC EAA3d (R63068-R63070) occur naturally in human brain. They are all
CC members of the EAA3 receptor family, having extracellular N- and C-
CC termini and 4 internal hydrophobic domains which anchor the receptor
CC within the cell surface membrane. The receptors bind glutamate and
CC also exhibit binding properties characteristic of kainate-type EAA
CC receptors.
CC N.B. The cDNA and corresp. amino acid sequences for EAA3b, 3c and
CC 3d do not appear in full in the specification; the sequences have
CC been compiled by combining sequences in Figures 1 and 4 according
CC to the description given in the disclosure.
CC Sequence 905 AA;
SQ

Query Match 60.5%; Score 52; DB 11; Length 905;
Best Local Similarity 60.0%; Pred. No. 3.71e+01;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 756 gglidskgyg 765
QY 70 GALVDSKSYA 79

RESULT 8
ID R60112 standard; Protein: 905 AA.
AC R60112.
DT 22-MAR-1995 (first entry)
DE Human EAA3a excitatory amino acid receptor.
KW excitatory amino acid receptor; human EAA3 receptor; kainate-type;
KW EAA3a receptor.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..30
FT /label= signal_peptide
FT protein 31..905
FT /label= mature_EAA3a_receptor
FT domain 31..562
FT /label= N-terminal_domain
FT /note= "extracellular"
FT region 563..840
FT /label= hydrophobic_region
FT domain 563..582
FT /label= TM-1
FT domain 604..624
FT /label= TM-2
FT /note= "transmembrane domain"
FT domain 635..653
FT /label= TM-3
FT /note= "transmembrane domain"
FT domain 820..840
FT /label= TM-4
FT /note= "transmembrane domain"
FT domain 841..905
FT /label= C-terminal_domain
FT /note= "extracellular"

```

```

PN CA2110933-A.
PD 12-JUN-1994.
PF 08-DEC-1993; 110933.
PR 11-DEC-1992; US-989793.
PA (KAMB/) KAMBOJ R.
PA (ELLI/) ELLIOT C E.
PA (NUTT/) NUTT S L.
PI Elliot CE, Kamboj R, Nutt SL;
DR WPI: 94-255829/32.
DR N-PSDB: Q70097.
PT Polynucleotide encoding a human excitatory amino acid 3 receptor
PT or fragment - used to assay test ligands for their interaction
PT with a human CBS receptor.
PS Claim 13; Fig 1; 35pp; English.
CC The sequence coding for the EAA3a receptor was isolated by probing a
CC human foetal brain cDNA library. The EAA3a receptor and naturally
CC occurring variants of it (i.e. EAA3b, EAA3c and EAA3d) are members
CC of the EAA3 receptor family; they all have extracellular N- and C-
CC termini and 4 internal hydrophobic domains which anchor the receptor
CC within the cell surface membrane. The receptors bind glutamate and
CC also exhibit binding properties characteristic of kainate-type EAA
CC receptors.
CC Sequence 905 AA;
SQ

Query Match 60.5%; Score 52; DB 11; Length 905;
Best Local Similarity 60.0%; Pred. No. 3.71e+01;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 756 gglidskgyg 765
QY 70 GALVDSKSYA 79

RESULT 9
ID R75884 standard; Protein: 908 AA.
AC R75884;
DT 10-JUL-1996 (first entry)
DE Human EAA4 receptor (I532, Y536, Q586).
KW Glutamate receptor; EAA4 receptor; excitatory amino acid;
KW CNS receptor; RNA editing.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1..31
FT /label= sig_peptide
FT protein 32..908
FT /label= Mat_protein
FT WO9517508-A2.
PD 29-JUN-1995.
PF 21-DEC-1994; CA0705.
PR 23-DEC-1993; US-172188.
PI (ALIX ) ALLELIX BIOPHARMACEUTICALS INC.
PI Kamboj R, Nutt S;
DR WPI: 95-240670/31.
DR N-PSDB: Q91232.
PT Identification of human CNS receptor ligand - and identification of
PT agents that modulate editing of human CNS receptors
PS Claim 5; Fig 6; 59pp; English.
CC The amino acid sequence (R75884) of human EAA4 receptor deduced from
CC a genomic DNA sequence (Q91232) differed from that deduced from
CC EAA4 cDNA; the genomic sequence coded for Ile, Tyr and Gln at
CC positions 532, 536 and 586, respectively, in the mature protein,
CC and the cDNA sequence for Val, Cys and Arg at these positions.
CC These changes were each a result of A to G substitutions in the
CC coding sequence and a consequence of RNA editing. The edited and
CC unedited receptors were functionally distinct and appeared to be
CC selectively expressed. CNS receptors subject to such editing (see
CC also R75882 and R75884) have value in screening cpds. for potential
CC therapeutic utility.
CC Sequence 908 AA;
SQ

Query Match 60.5%; Score 52; DB 16; Length 908;
Best Local Similarity 60.0%; Pred. No. 3.71e+01;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

Db 756 gglidskgyg 765
 QY 70 GALVDSKSYA 79

RESULT 10
 ID R45367 standard; Protein: 908 AA.
 AC R45367;
 DT 07-JUL-1994 (first entry)
 DE Excitatory amino acid receptor (EAA4a receptor).
 KW Excitatory amino acid receptor; EAA4a; screening; detection;
 KW identification; CNS; nervous disorders; antibody; oocyte; membrane
 protein; ion-channel; therapeutics.
 OS Homo sapiens.
 PN EP-578409-A.
 PD 12-JAN-1994.
 PF 24-JUN-1993; 304961.
 PR 24-JUN-1992; US-903456.
 PA (ELLIOTT) ELLIOTT C E.
 PA (KAMB/) KAMBOJ R.
 PA (NUTT/) NUTT S L.
 PI Kamboj R, Nutt SL, Elliott CE;
 DR WPI: 94-010248/02.
 DR N-PSDB; Q54482.
 PT Excitatory amino acid for receptors with kainate binding activity
 PT - used to test ligands for CNS receptor interaction to identify
 PT cpds. useful against CNS disorders
 PS Claim 13; Figure 1; 37pp; English.
 CC The excitatory amino acid receptor can be used to screen ligands.
 CC The ligands can be assayed for interaction by incubating them with
 CC cells, preferably oocytes, expressing the receptor or with membrane
 CC preparations from these cells, and assessing any interaction by
 CC determining receptor-ligand binding or ligand-mediated ion channel
 CC activation. This method is useful in the identification of
 CC therapeutics useful to treat CNS disorders in humans. Fragments of
 CC the receptor are useful for structural investigations and to raise
 CC antibodies.
 SQ Sequence 908 AA;

Query Match 60.5%; Score 52; DB 9; Length 908;
 Best Local Similarity 60.0%; Pred. No. 3.71e+01;
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 756 gglidskgyg 765
 QY 70 GALVDSKSYA 79

RESULT 11
 ID R50179 standard; Protein: 919 AA.
 AC R50179;
 DT 17-OCT-1994 (first entry)
 DE Excitatory amino acid receptor (EAA5).
 KW Excitatory amino acid receptor; EAA; identification; detection;
 KW CNS; central nervous system; therapeutic; antibody; ligand;
 KW screening.
 OS Homo sapiens.
 PN EP-588642-A.
 PD 23-MAR-1994.
 PF 16-SEP-1993; 307325.
 PR 17-SEP-1992; US-945210.
 PA (ELLIOTT) ELLIOTT C E.
 PA (KAMB/) KAMBOJ R.
 PA (NUTT/) NUTT S L.
 PI Elliott CE, Kamboj R, Nutt SL;
 DR WPI: 94-094202/12.
 DR N-PSDB; Q44685.
 PT Nucleic acid encoding human excitatory amino acid receptors - used
 PT for producing receptors and in assays for test ligands for
 PT binding to human CNS receptors
 PS Claim 12; page 16-19; 34pp; English.
 CC Nucleotides which encode the excitatory amino acid receptors (EAA's)

CC can be used for the production of the receptors, to identify
 CC sequence related genes or for locating the receptor encoding DNA in
 CC a specimen. Recombinant cells which produce the receptors, or their
 CC membrane preparations, can be used for assaying a test ligand for
 CC binding to a human CNS receptor to develop therapeutics. The
 CC receptors can themselves be used in the production of antibodies for
 CC use in detection methods.
 SQ Sequence 919 AA;

Query Match 60.5%; Score 52; DB 10; Length 919;
 Best Local Similarity 60.0%; Pred. No. 3.71e+01;
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 757 gglidskgyg 766
 QY 70 GALVDSKSYA 79

RESULT 12
 ID R11993 standard; Protein: 920 AA.
 AC R11993;
 DT 31-JUL-1991 (first entry)
 DE Glutamate receptor 5-1.
 KW Glutamate receptor 5-1; probe; ligand; drug screening.
 OS Rattus rattus.
 FH Key Location/Qualifiers
 FT peptide 1..30
 FT /label= sig_peptide
 FT protein 31..920
 FT /label= mat_protein
 FT region 402..416
 FT /label= insertion
 PN W09106648-A.
 PD 16-MAY-1991.
 PF 25-OCT-1990; U06153.
 PR 27-OCT-1989; US-428116.
 PA (SALK) SALK INST FOR BIOL STUD.
 PI Heinemann SF, Boulter JR, Hollmann M, Bettler B, Jensen JE;
 DR WPI: 91-164197/22.
 DR N-PSDB; Q11853.
 PT Glutamate receptors - used to screen for functional ligands and
 PT identify and isolate further receptors
 PS Disclosure; Fig 10; 109pp; English.
 CC GluR5-1 has a 15 amino acid insert (see features) compared to the
 CC shorter variant GluR5-2 and is unique among the receptors GluR1-7.
 CC It has a Mr of 100,000. The signal sequence cleavage site is after
 CC a Pro, which is atypical.
 CC The gene and protein can be used in drug screening, to
 CC determine whether a substance is a functional ligand for the
 CC receptor by monitoring ion channel activity.
 CC See also Q11849-855.
 SQ Sequence 920 AA;

Query Match 60.5%; Score 52; DB 2; Length 920;
 Best Local Similarity 60.0%; Pred. No. 3.71e+01;
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 771 gglidskgyg 780
 QY 70 GALVDSKSYA 79

RESULT 13
 ID R45369 standard; Protein: 50 AA.
 AC R45369;
 DT 07-JUL-1994 (first entry)
 DE Excitatory amino acid receptor (EAA4b) fragment.
 KW Excitatory amino acid receptor; EAA4a; screening; detection;
 KW identification; CNS; nervous disorders; antibody; oocyte; membrane
 KW protein; ion-channel; therapeutics.
 OS Homo sapiens.
 PN EP-578409-A.
 PD 12-JAN-1994.

PF 24-JUN-1993; 304961.
 PR 24-JUN-1992; US-903456.
 PA (ELLI/) ELLIOTT C E.
 PA (KAMB/) KAMBOJ R.
 PA (NUTT/) NUTT S L.
 PI Kamboj R, Nutt SL, Elliott CE;
 DR WPI: 94-010248/02.
 PT Excitatory aminoacid for receptors with kainate binding activity
 PT - used to test ligands for CNS receptor interaction to identify
 PT cpds. useful against CNS disorders
 PS Disclosure: Figure 3a; 37pp; English.
 CC The excitatory amino acid receptor EAA4a can be used to screen
 CC ligands. The ligands can be assayed for interaction by incubating
 CC them with cells, preferably oocytes, expressing the receptor or with
 CC membrane preparations from these cells, and assessing any
 CC in-eraction by determining receptor-ligand binding or
 CC ligand-mediated ion channel activation. This method is useful in
 CC the identification of therapeutics useful to treat CNS disorders in
 CC humans. Fragments of the receptor are useful for structural
 CC investigations and to raise antibodies. This fragment of the
 CC receptor EAA4b corresponds to amino acid residues 700 to 750 and has
 CC aspartic acid at position 727 as opposed to glycine which occurs in
 CC the corresponding position in receptor-EAA4a.
 SQ Sequence 50 AA;

Query Match 59.3%; Score 51; DB 9; Length 50;
 Best Local Similarity 60.0%; Pred. No. 4.90e+01;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 26 gdlidskgvg 35
 | :|:|:|:
 QY 70 GALVDSKSYA 79

RESULT 14

ID P60011 standard; Protein; 259 AA.
 AC P60011;
 DT 30-JUL-1991 (first entry)
 DE Sequence of oncogene v-fms.
 KW Therapy: diagnosis; insulin.
 PN EP-192392-A.
 PD 27-AUG-1986.
 PF 10-FEB-1986; 300894.
 PR 12-FEB-1985; US-700776.
 PA (GETH) GENENTECH INC.
 PI Bell JR, Ulrich A, Ramachandran J;
 DR WPI: 86-226966/35.
 PT New DNA encoding insulin receptor or its fragments - used for
 PT synthesis of receptor and mutants for therapeutic and diagnostic
 PT use
 PS Example: Fig 5; 62pp; English.
 CC A mutant IR is claimed which may have a mutated alpha-chain, esp. at
 CC the precursor processing site. The beta-chain may be mutated, e.g.
 CC by deletion of the transmembrane sequence; the tyrosine kinase
 CC activity may be inactivated. Fig. 5 is a comparison of oncogene and
 CC human EGF receptor sequences with that of HIR in the cytoplasmic
 CC domain of the insulin receptor beta subunit.
 SQ Sequence 259 AA;

Query Match 58.1%; Score 50; DB 3; Length 259;
 Best Local Similarity 53.8%; Pred. No. 6.45e+01;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 191 npypgilvnskyf 203
 : | | | | |
 QY 66 DPSTGALVDSKSY 78

RESULT 15

ID W41513 standard; Protein; 371 AA.
 AC W41513;
 DT 17-JUN-1998 (first entry)
 DE N. meningitidis alpha-2,3-sialyltransferase protein.

KW Alpha-2,3-sialyltransferase; sialic acid; acceptor; lipid;
 KW biologically active oligosaccharide; sialyl-modified protein.
 OS Neisseria meningitidis.
 PN WO9747749-A1.
 PD 18-DEC-1997.

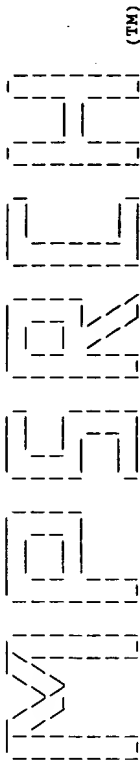
PF 10-JUN-1997; CA0390.
 PR 06-JUN-1997; US-872485.
 PR 10-JUN-1996; US-019520.
 PA (CANA) NAT RES COUNCIL CANADA.
 PI Gilbert M, Jennings MP, Wakarchuk WW, Young NM;
 DR WPI: 98-052313/05.
 DR N-PSDB; V04125.

PT Nucleic acid sequence encoding Neisseria
 PT alpha-2,3-sialyltransferase - useful to add sialic acid to acceptor
 PT with terminal galactose residue for synthesis of biologically active
 PT oligosaccharide.
 PS Claim 31; Page 38; 50pp; English.
 CC This sequence represents an alpha-2,3-sialyltransferase which has been
 CC isolated from Neisseria meningitidis. The protein can be used as a
 CC reagent for adding a sialic acid residue to an acceptor having a
 CC terminal galactose residue, e.g. in synthesis of biologically active
 CC oligosaccharides or sialyl-modified proteins or lipids. The nucleic
 CC acid sequence can be used for the recombinant production of
 CC alpha-2,3-sialyltransferase.
 SQ Sequence 371 AA;

Query Match 58.1%; Score 50; DB 29; Length 371;
 Best Local Similarity 53.8%; Pred. No. 6.45e+01;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 164 ddgtgnliqssy 176
 | :| | :|:
 QY 66 DPSTGALVDSKSY 78

Search completed: Wed Sep 1 16:30:56 1999
 Job time : 25 secs.



Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Sep 1 16:30:02 1999; MasPar time 3.68 Seconds
Tabular output not generated. 152.280 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (66-79) from PCTUS9913024.ppt (11 of 12)
Perfect Score: 86
Sequence: 1 DPSTGALVDSKSYA 14

Scoring table: PAM 150
Gap 11
Searched: 122810 seqs, 40068593 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4
Statistics: Mean 24.770; Variance 28.525; scale 0.868

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	52	60.5	864	2 A43954	glutamate receptor be	4.06e+00
2	52	60.5	869	2 S35792	glutamate receptor G1	4.06e+00
3	52	60.5	908	2 A54260	glutamate receptor 6	4.06e+00
4	52	60.5	908	2 S19098	glutamate receptor G1	4.06e+00
5	52	60.5	918	2 I58178	glutamate receptor -	4.06e+00
6	52	60.5	919	2 S19810	glutamate receptor G1	4.06e+00
7	52	60.5	919	2 I53474	kainate receptor - ra	4.06e+00
8	52	60.5	949	2 S19808	glutamate receptor G1	4.06e+00
9	51	59.3	469	2 S17813	crtJ protein - rhodob	6.60e+00
10	51	58.3	563	2 S04886	mutC protein - Vibrio	6.60e+00
11	50	58.1	331	2 A28063	glutamin-(asparagin-)	1.06e+01
12	50	58.1	542	2 JC5507	monocarboxylate trans	1.06e+01
13	50	58.1	941	1 TVNVMQ	protein-tyrosine kina	1.06e+01
14	50	58.1	972	1 TVHUMD	macrophage colony-sti	1.06e+01
15	50	58.1	980	1 TVCTMD	macrophage colony-sti	1.06e+01
16	50	58.1	1133	2 T01757	hypothetical protein	1.70e+01
17	49	57.0	235	2 S48382	hypothetical protein	1.70e+01
18	49	57.0	319	2 A69756	adhesion protein homo	1.70e+01
19	49	57.0	357	2 A54512	serine-repeat antigen	1.70e+01
20	49	57.0	427	2 A34639	parasitophorous vacuo	1.70e+01
21	49	57.0	477	1 VGBEG1	glycoprotein E - suid	1.70e+01
22	49	57.0	635	1 SCYHAL	S-aminolevulinic syn	1.70e+01
23	49	57.0	966	2 G69189	hypothetical protein	1.70e+01

24	49	57.0	989	2 A54505	serine-repeat antigen	1.70e+01
25	49	57.0	997	2 B71617	SERA antigen/papain-1	1.70e+01
26	48	55.8	195	2 E70555	hypothetical protein	2.70e+01
27	48	55.8	427	2 S11889	S-locus-specific glyco	2.70e+01
28	48	55.8	434	2 B33223	SLR2 protein (S6 alle	2.70e+01
29	48	55.8	439	2 A53223	SLR2 protein (S2 alle	2.70e+01
30	48	55.8	448	2 P50047	serine proteinase (EC	2.70e+01
31	48	55.8	676	2 S69783	outer membrane protei	2.70e+01
32	48	55.8	683	2 S69780	outer membrane protei	2.70e+01
33	48	55.8	707	2 S69781	outer membrane protei	2.70e+01
34	48	55.8	712	2 S69782	outer membrane protei	2.70e+01
35	47	54.7	232	2 A71068	hypothetical protein	4.25e+01
36	47	54.7	257	2 F70166	hypothetical protein	4.25e+01
37	47	54.7	398	2 B71284	probable periplasmic	4.25e+01
38	47	54.7	432	2 S67821	GumE protein - xantho	4.25e+01
39	47	54.7	434	2 S04534	Invasin precursor - Y	4.25e+01
40	47	54.7	501	2 S35583	splicing factor SF3a6	4.25e+01
41	47	54.7	501	2 A55749	spliceosome-associate	4.25e+01
42	47	54.7	503	1 YF8YAC	phenylalanine--tRNA	4.25e+01
43	47	54.7	606	2 A70960	probable pckA protein	4.25e+01
44	47	54.7	747	2 S37694	gene PC326 protein -	4.25e+01
45	47	54.7	2233	1 ZLNZP3	genome polyprotein -	4.25e+01

ALIGNMENTS

RESULT 1
ENTRY A43954 #type complete
TITLE glutamate receptor beta-2 chain - mouse
ALTERNATE_NAMES kainate receptor
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 06-Sep-1996
ACCESSIONS A43954; I46131
REFERENCE A43954
#authors Morita, T.; Sakimura, K.; Kushiya, E.; Yamazaki, M.; Meguro, H.; Araki, K.; Abe, T.; Mori, K.J.; Mishina, M.
#journal Brain Res. Mol. Brain Res. (1992) 14:143-146
#title Cloning and functional expression of a cDNA encoding the mouse beta 2 subunit of the kainate-selective glutamate receptor channel.
#cross-references MUID:92356790
#accession A43954
#status nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-864 #label MOR
#cross-references GB:D10054; NID:g220408; PID:g220409
#experimental_source forebrain
#note #sequence extracted from NCBI backbone (NCBIP:110532)

GENETICS
#gene pGluRbeta-2
CLASSIFICATION #superfamily glutamate receptor; glutamate receptor homology
KEYWORDS neurotransmitter receptor; transmembrane protein
FEATURE
#domain glutamate receptor homology #label GRH
#length 864 #molecular-weight 97353 #checksum 9801
SUMMARY

Query Match 60.5%; Score 52; DB 2; Length 864;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 756 GGLIDSKGYG 765
QY 70 GALVDSKSYA 79

RESULT 2
ENTRY S35792 #type complete
TITLE glutamate receptor GluR6C - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Sep-1997
ACCESSIONS S35792

```

REFERENCE S35792
#authors Gregor, P.
#submission submitted to the EMBL Data Library, June 1992
#accession S35792
##status preliminary
##molecule_type mRNA
##residues 1-869 ##label GRE
##cross-references EMBL:X56117; NID:g312493; PID:g312494
CLASSIFICATION #superfamily glutamate receptor; glutamate receptor homology
KEYWORDS neurotransmitter receptor; transmembrane protein
FEATURE
435-855 #domain glutamate receptor homology #label GRH
SUMMARY #length 869 #molecular-weight 97807 #checksum 1988

Query Match 60.5%; Score 52; DB 2; Length 869;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 756 GGLIDSKGYG 765
QY 70 GALVDSKSYA 79

RESULT 3
ENTRY A54260 #type complete
TITLE glutamate receptor 6 kainate-preferring precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 03-Sep-1994 #sequence_revision 09-Sep-1994 #text_change
29-Aug-1997
ACCESSIONS A54260
REFERENCE
#authors Paschen, W.; Blackstone, C.D.; Huganir, R.L.; Ross, C.A.
#journal Genomics (1994) 20:435-440
#title Human GluR6 kainate receptor (GRIK2): molecular cloning,
expression, polymorphism, and chromosomal assignment.
#cross-references MUID:94307730
#accession A54260
##status preliminary; not compared with conceptual translation
##molecule_type mRNA
##residues 1-908 ##label PAS
GENETICS
#gene GDB:GRIK2
##cross-references GDB:l31461; OMIM:138244
#map_position 6q16.3-q21
CLASSIFICATION #superfamily glutamate receptor; glutamate receptor homology
KEYWORDS neurotransmitter receptor
FEATURE
435-855 #domain glutamate receptor homology #label GRH
SUMMARY #length 908 #molecular-weight 102612 #checksum 2705

Query Match 60.5%; Score 52; DB 2; Length 908;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 756 GGLIDSKGYG 765
QY 70 GALVDSKSYA 79

RESULT 4
ENTRY S19098 #type complete
TITLE glutamate receptor GluR6 precursor - rat
ALTERNATE_NAMES kainate receptor
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change
08-Sep-1997
ACCESSIONS S19098; S16234; S19809
REFERENCE
#authors Bettler, B.
#submission submitted to the EMBL Data Library, December 1991
#accession S19098
##molecule_type mRNA
##residues 1-908 ##label BET

```

```

##cross-references EMBL:Z11548; NID:g56281; PID:g56282
REFERENCE S16234
#authors Egebjerg, J.; Bettler, B.; Hermans-Borgmeyer, I.; Heinemann,
S.
#journal Nature (1991) 351:745-748
#title Cloning of a cDNA for a glutamate receptor subunit activated
by kainate but not AMPA.
#cross-references MUID:91287799
#accession S16234
#status nucleic acid sequence not shown
##molecule_type mRNA
##residues 1-880,'POLL' ##label EGE
##cross-references EMBL:Z11548
REFERENCE S19806
#authors Seeburg, P.H.
#submission submitted to the EMBL Data Library, February 1992
#accession S19809
##molecule_type mRNA
##residues 1-620,'Q',622-908 ##label SEE
##cross-references EMBL:Z11715; NID:g56279; PID:g56280
CLASSIFICATION #superfamily glutamate receptor; glutamate receptor homology
KEYWORDS neurotransmitter receptor; transmembrane protein
FEATURE
1-31 #domain signal sequence #status predicted #label SIG\
32-908 #product glutamate receptor GluR6 #status predicted
#label MAR\
435-855 #domain glutamate receptor homology #label GRH\
505-521 #domain transmembrane #status predicted #label TM1\
563-581 #domain transmembrane #status predicted #label TM2\
635-653 #domain transmembrane #status predicted #label TM3\
820-840 #domain transmembrane #status predicted #label TM4
SUMMARY #length 908 #molecular-weight 102423 #checksum 3737

Query Match 60.5%; Score 52; DB 2; Length 908;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 756 GGLIDSKGYG 765
QY 70 GALVDSKSYA 79

RESULT 5
ENTRY I58178 #type complete
TITLE glutamate receptor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
28-Aug-1998
ACCESSIONS I58178
REFERENCE
#authors Gregor, P.; O'Hara, B.F.; Yang, X.; Uhl, G.R.
#journal Neuroreport (1993) 4:1343-1346
#title Expression and novel subunit isoforms of glutamate receptor
genes GluR5 and GluR6.
#cross-references MUID:94083547
#accession I58178
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1-918 ##label RES
##cross-references GB:L19058; NID:g455447; PID:g455448
GENETICS
#gene GDB:GRIK1; GLUR5
##cross-references GDB:l31462; OMIM:138245
#map_position 21q22.1-21q22.2
CLASSIFICATION #superfamily glutamate receptor; glutamate receptor homology
KEYWORDS neurotransmitter receptor
SUMMARY #length 918 #molecular-weight 103980 #checksum 1748

Query Match 60.5%; Score 52; DB 2; Length 918;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 771 GGLIDSKGYG 780

```

```

QY      70 GALVDSKSYA 79

RESULT      6
ENTRY       S19810          #type complete
TITLE       glutamate receptor GluR7 - rat
ALTERNATE_NAMES
ORGANISM    #formal_name Rattus norvegicus #common_name Norway rat
DATE        30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
            08-Sep-1997
ACCESSIONS  S19810; JH0587
REFERENCE    Seeburg, P.H.
             submitted to the EMBL Data Library, February 1992
             #submission
             #accession
             ##molecule_type mRNA
             ##residues 1-919 ##label SEE
             ##cross-references EMBL:Z11716
REFERENCE    Bettler, B.; Egebjerg, J.; Sharma, G.; Pecht, G.;
             Hermans-Borgmeyer, I.; Moll, C.; Stevens, C.F.; Heinemann,
             S.
             Neuron (1992) 8:257-265
             Cloning of a putative glutamate receptor: a low affinity
             kainate-binding subunit.
             #cross-references MUID:92153417
             #accession JH0587
             ##status
             ##molecule_type mRNA
             ##residues 32-117,'I','L19-175','T','177-209','D','211','DD','214-289','P',
             291-340,'P','342-354','A','356-459','DR','462-466','ID',
             469-919 ##label BET
             ##cross-references GB:M83552; NID:g204391; PID:g204392
             ##experimental_source cerebellum
COMMENT      Glutamate receptors form part of a major excitatory
             neurotransmitter system of the brain and play roles in synapse
             formation, in learning and memory, and in the pathogenesis of
             neurodegenerative diseases.
CLASSIFICATION #superfamily glutamate receptor; glutamate receptor homology
KEYWORDS       glycoprotein; neurotransmitter receptor; phosphoprotein;
             transmembrane protein
FEATURE        438-856   #domain glutamate receptor homology #label GRH\
507-524   #domain transmembrane #status predicted #label TM1\
565-584   #domain transmembrane #status predicted #label TM2\
608-626   #domain transmembrane #status predicted #label TM3\
637-663   #domain transmembrane #status predicted #label TM4\
821-841   #domain transmembrane #status predicted #label TM5\
415         #binding_site carbohydrate (Asn) (covalent) #status
              predicted\
668         #binding_site phosphate (Ser) (covalent) #status
              predicted
SUMMARY        #length 919 #molecular-weight 103991 #checksum 9539
Query Match    60.5% Score 52; DB 2; Length 919;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 757 GGLIDSKGYG 766
|::|||::|
QY      70 GALVDSKSYA 79

RESULT      7
ENTRY       I53474          #type complete
TITLE       kainate receptor - rat
ALTERNATE_NAMES
ORGANISM    #formal_name Rattus norvegicus #common_name Norway rat
DATE        02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change
            16-Feb-1997
ACCESSIONS  I53474
REFERENCE    Lomeli, H.; Wisden, W.; Kohler, M.; Keinonen, K.; Sommer, B.;
             Seeburg, P.H.
             submitted to the EMBL Data Library, February 1992
             #submission
             #accession
             ##molecule_type mRNA
             ##residues 1-919 ##label SEE
             ##cross-references EMBL:Z11716
REFERENCE    Bettler, B.; Egebjerg, J.; Sharma, G.; Pecht, G.;
             Hermans-Borgmeyer, I.; Moll, C.; Stevens, C.F.; Heinemann,
             S.
             Neuron (1992) 8:257-265
             Cloning of a putative glutamate receptor: a low affinity
             kainate-binding subunit.
             #cross-references MUID:92153417
             #accession JH0587
             ##status
             ##molecule_type mRNA
             ##residues 32-117,'I','L19-175','T','177-209','D','211','DD','214-289','P',
             291-340,'P','342-354','A','356-459','DR','462-466','ID',
             469-919 ##label BET
             ##cross-references GB:M83552; NID:g204391; PID:g204392
             ##experimental_source cerebellum
COMMENT      Glutamate receptors form part of a major excitatory
             neurotransmitter system of the brain and play roles in synapse
             formation, in learning and memory, and in the pathogenesis of
             neurodegenerative diseases.
CLASSIFICATION #superfamily glutamate receptor; glutamate receptor homology
KEYWORDS       glycoprotein; neurotransmitter receptor; phosphoprotein;
             transmembrane protein
FEATURE        438-856   #domain glutamate receptor homology #label GRH\
507-524   #domain transmembrane #status predicted #label TM1\
565-584   #domain transmembrane #status predicted #label TM2\
608-626   #domain transmembrane #status predicted #label TM3\
637-663   #domain transmembrane #status predicted #label TM4\
821-841   #domain transmembrane #status predicted #label TM5\
415         #binding_site carbohydrate (Asn) (covalent) #status
              predicted\
668         #binding_site phosphate (Ser) (covalent) #status
              predicted
SUMMARY        #length 919 #molecular-weight 103991 #checksum 9539
Query Match    60.5% Score 52; DB 2; Length 919;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 757 GGLIDSKGYG 766
|::|||::|
QY      70 GALVDSKSYA 79

RESULT      7
ENTRY       I53474          #type complete
TITLE       kainate receptor - rat
ALTERNATE_NAMES
ORGANISM    #formal_name Rattus norvegicus #common_name Norway rat
DATE        02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change
            16-Feb-1997
ACCESSIONS  I53474
REFERENCE    Lomeli, H.; Wisden, W.; Kohler, M.; Keinonen, K.; Sommer, B.;
             Seeburg, P.H.
             submitted to the EMBL Data Library, February 1992
             #submission
             #accession
             ##molecule_type mRNA
             ##residues 1-919 ##label SEE
             ##cross-references EMBL:Z11716
REFERENCE    Bettler, B.; Egebjerg, J.; Sharma, G.; Pecht, G.;
             Hermans-Borgmeyer, I.; Moll, C.; Stevens, C.F.; Heinemann,
             S.
             Neuron (1992) 8:257-265
             Cloning of a putative glutamate receptor: a low affinity
             kainate-binding subunit.
             #cross-references MUID:92153417
             #accession JH0587
             ##status
             ##molecule_type mRNA
             ##residues 32-117,'I','L19-175','T','177-209','D','211','DD','214-289','P',
             291-340,'P','342-354','A','356-459','DR','462-466','ID',
             469-919 ##label BET
             ##cross-references GB:M83552; NID:g204391; PID:g204392
             ##experimental_source cerebellum
COMMENT      Glutamate receptors form part of a major excitatory
             neurotransmitter system of the brain and play roles in synapse
             formation, in learning and memory, and in the pathogenesis of
             neurodegenerative diseases.
CLASSIFICATION #superfamily glutamate receptor; glutamate receptor homology
KEYWORDS       glycoprotein; neurotransmitter receptor; phosphoprotein;
             transmembrane protein
FEATURE        438-856   #domain glutamate receptor homology #label GRH\
507-524   #domain transmembrane #status predicted #label TM1\
565-584   #domain transmembrane #status predicted #label TM2\
608-626   #domain transmembrane #status predicted #label TM3\
637-663   #domain transmembrane #status predicted #label TM4\
821-841   #domain transmembrane #status predicted #label TM5\
415         #binding_site carbohydrate (Asn) (covalent) #status
              predicted\
668         #binding_site phosphate (Ser) (covalent) #status
              predicted
SUMMARY        #length 919 #molecular-weight 103991 #checksum 9539
Query Match    60.5% Score 52; DB 2; Length 919;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 757 GGLIDSKGYG 766
|::|||::|
QY      70 GALVDSKSYA 79

RESULT      7
ENTRY       I53474          #type complete
TITLE       kainate receptor - rat
ALTERNATE_NAMES
ORGANISM    #formal_name Rattus norvegicus #common_name Norway rat
DATE        02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change
            16-Feb-1997
ACCESSIONS  I53474
REFERENCE    Lomeli, H.; Wisden, W.; Kohler, M.; Keinonen, K.; Sommer, B.;
             Seeburg, P.H.
             submitted to the EMBL Data Library, February 1992
             #submission
             #accession
             ##molecule_type mRNA
             ##residues 1-919 ##label SEE
             ##cross-references EMBL:Z11716
REFERENCE    Bettler, B.; Egebjerg, J.; Sharma, G.; Pecht, G.;
             Hermans-Borgmeyer, I.; Moll, C.; Stevens, C.F.; Heinemann,
             S.
             Neuron (1992) 8:257-265
             Cloning of a putative glutamate receptor: a low affinity
             kainate-binding subunit.
             #cross-references MUID:92153417
             #accession JH0587
             ##status
             ##molecule_type mRNA
             ##residues 32-117,'I','L19-175','T','177-209','D','211','DD','214-289','P',
             291-340,'P','342-354','A','356-459','DR','462-466','ID',
             469-919 ##label BET
             ##cross-references GB:M83552; NID:g204391; PID:g204392
             ##experimental_source cerebellum
COMMENT      Glutamate receptors form part of a major excitatory
             neurotransmitter system of the brain and play roles in synapse
             formation, in learning and memory, and in the pathogenesis of
             neurodegenerative diseases.
CLASSIFICATION #superfamily glutamate receptor; glutamate receptor homology
KEYWORDS       glycoprotein; neurotransmitter receptor; phosphoprotein;
             transmembrane protein
FEATURE        438-856   #domain glutamate receptor homology #label GRH\
507-524   #domain transmembrane #status predicted #label TM1\
565-584   #domain transmembrane #status predicted #label TM2\
608-626   #domain transmembrane #status predicted #label TM3\
637-663   #domain transmembrane #status predicted #label TM4\
821-841   #domain transmembrane #status predicted #label TM5\
415         #binding_site carbohydrate (Asn) (covalent) #status
              predicted\
668         #binding_site phosphate (Ser) (covalent) #status
              predicted
SUMMARY        #length 919 #molecular-weight 103991 #checksum 9539
Query Match    60.5% Score 52; DB 2; Length 919;
Best Local Similarity 60.0%; Pred. No. 4.06e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 757 GGLIDSKGYG 766
|::|||::|
QY      70 GALVDSKSYA 79

RESULT      7
ENTRY       I53474          #type complete
TITLE       kainate receptor - rat
ALTERNATE_NAMES
ORGANISM    #formal_name Rattus norvegicus #common_name Norway rat
DATE        02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change
            16-Feb-1997
ACCESSIONS  I53474
REFERENCE    Lomeli, H.; Wisden, W.; Kohler, M.; Keinonen, K.; Sommer, B.;
             Seeburg, P.H.
             submitted to the EMBL Data Library, February 1992
             #submission
             #accession
             ##molecule_type mRNA
             ##residues 1-919 ##label SEE
             ##cross-references EMBL:Z11716
REFERENCE    Bettler, B.; Egebjerg, J.; Sharma, G.; Pecht, G.;
             Hermans-Borgmeyer, I.; Moll, C.; Stevens, C.F.; Heinemann,
             S.
             Neuron (1992) 8:257-265
             Cloning of a putative glutamate receptor: a low affinity
             kainate-binding subunit.
             #cross-references MUID:92153417
             #accession JH0587
             ##status
             ##molecule_type mRNA
             ##residues 32-117,'I','L19-175','T','177-209','D','211','DD','214-289','P',
             291-340,'P','342-354','A','356-459','DR','462-466','ID',
             469-919 ##label BET
             ##cross-references GB:M83552; NID:g204391; PID:g204392
             ##experimental_source cerebellum
COMMENT      Glutamate receptors form part of a major excitatory
```

```

Db 771 GGLIDSKGYG 780
   1-1-1111-1-1
QY 70 GALVDSKSYA 79

RESULT 9
ENTRY crtJ protein - Rhodobacter capsulatus
TITLE #formal_name Rhodobacter capsulatus
ORGANISM 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
DATE 30-Jun-1992

ACCESSIONS S17813 #type complete
REFERENCE S17803
AUTHORS Burke, D.H.; Alberti, M.; Armstrong, G.A.; Hearst, J.E.
SUBMISSION submitted to the EMBL Data Library, November 1991
DESCRIPTION The complete nucleotide sequence of the 46 kb photosynthesis
#accession S17813 gene cluster of Rhodobacter capsulatus.
#molecule_type DNA
#residues 1-469 #label EMB
#cross-references EMBL:Z11165
GENETICS
#gene crtJ
SUMMARY #length 469 #molecular-weight 51331 #checksum 8204
Query Match 59.3%; Score 51; DB 2; Length 469;
Best Local Similarity 42.9%; Pred. No. 6.60e+00;
Matches 6; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Db 308 DAGSAALVQGRSFA 321
   1-1-1111111-1-1
QY 66 DPSTGALVDSKSYA 79

RESULT 10
ENTRY S04886 #type complete
TITLE mutL protein - Vibrio cholerae
ORGANISM #formal_name Vibrio cholerae
DATE 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change
09-Sep-1997

ACCESSIONS S04886
REFERENCE S04886
AUTHORS Bera, T.K.; Ghosh, S.K.; Das, J.
JOURNAL Nucleic Acids Res. (1989) 17:6241-6251
TITLE Cloning and characterization of mutL and mutS genes of Vibrio
#cross-references MUID:89366654
#accession S04886
#molecule_type DNA
#residues 1-563 #label BER
#cross-references EMBL:X15438; NID:g48368; PID:g48369
GENETICS
#gene mutL
SUMMARY #length 563 #molecular-weight 64858 #checksum 6294
Query Match 59.3%; Score 51; DB 2; Length 563;
Best Local Similarity 63.6%; Pred. No. 6.60e+00;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 293 DPTGTGLVERK 303
   1-1-1111-1-1
QY 66 DPSTGALVDSK 76

RESULT 11
ENTRY A28063 #type complete
TITLE glutamin-(asparagin-)ase (EC 3.5.1.38) - Acinetobacter
ALTERNATE_NAMES calcoaceticus
ORGANISM glutaminase-asparaginase
#formal_name Acinetobacter calcoaceticus
DATE 28-Aug-1989 #sequence_revision 28-Aug-1989 #text_change
28-Apr-1993

ACCESSIONS A28063
REFERENCE Tanaka, S.; Robinson, E.A.; Appella, E.; Miller, M.; Ammon,
H.L.; Roberts, J.; Weber, I.T.; Wlodawer, A.
J. Biol. Chem. (1988) 263:8583-8591
#journal Structures of amidohydrolases. Amino acid sequence of a
#title glutaminase-asparaginase from Acinetobacter
glutaminasificans and preliminary crystallographic data for
an asparaginase from Erwinia chrysanthemi.
#cross-references MUID:88243706
#accession A28063
#molecule_type protein
#residues 1-331 #label TAN
#note the source is designated as Acinetobacter
glutaminasificans
CLASSIFICATION #superfamily asparaginase
KEYWORDS hydrolase
SUMMARY #length 331 #molecular-weight 35485 #checksum 4526
Query Match 58.1%; Score 50; DB 2; Length 331;
Best Local Similarity 55.6%; Pred. No. 1.06e+01;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 182 GTLVEGKPY 190
   1-1-1111-1-1
QY 70 GALVDSKSY 78

RESULT 12
ENTRY JC5507 #type complete
TITLE monocarboxylate transporter 3 - chicken
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change
18-Sep-1998

ACCESSIONS JC5507
REFERENCE JC5507
AUTHORS Yoon, H.; Fanelli, A.; Grollman, E.F.; Philp, N.J.
JOURNAL Biochem. Biophys. Res. Commun. (1997) 234:90-94
#title Identification of a unique monocarboxylate transporter (MCT3)
in retinal pigment epithelium.
#cross-references MUID:97312526
#accession JC5507
#status nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-542 #label YOO
#cross-references GB:U15685
#experimental_source retinal pigment epithelial cell
COMMENT This protein regulates lactate levels in the interphotoreceptor
space.
CLASSIFICATION #superfamily monocarboxylate transporter MCT2
KEYWORDS phosphoprotein
FEATURE
141 #binding_site phosphate (Thr) (covalent) (by casein
kinase II) #status predicted\
210 #binding_site phosphate (Thr) (covalent) (by CAMP- and
cGMP-dependent kinases) #status predicted\
234,467 #binding_site phosphate (Ser) (covalent) (by casein
kinase II) #status predicted\
262,526 #binding_site phosphate (Ser) (covalent) (by protein
kinase C) #status predicted\
SUMMARY #length 542 #molecular-weight 58085 #checksum 7501
Query Match 58.1%; Score 50; DB 2; Length 542;
Best Local Similarity 61.5%; Pred. No. 1.06e+01;
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 410 PSAGRLVDALKNY 422
   1-1-1111-1-1
QY 67 PSTGALVDS-KSY 78

RESULT 13
ENTRY TVMVMD #type complete

```

```

TITLE      protein-tyrosine kinase (EC 2.7.1.112) fms precursor - feline
            sarcoma virus (strain McDonough)
ORGANISM   host Felis sp. (cat)
#note      27-Nov-1985 #sequence_revision 31-Dec-1991 #text_change
DATE       13-Jun-1997

ACCESSIONS A00654
REFERENCE   Hampe, A.; Gobet, M.; Sherr, C.J.; Galibert, F.
#authors    Proc. Natl. Acad. Sci. U.S.A. (1984) 81:85-89
#journal     Nucleotide sequence of the feline retroviral oncogene v-fms
#title       shows unexpected homology with oncogenes encoding
            tyrosine-specific protein kinases.
#cross-references MIM:84119469
#accession   A00654
#molecule_type DNA
#residues    1-941 #label HAM
COMMENT     This protein is synthesized as a gag-fms polyprotein.
GENETICS

CLASSIFICATION
#gene        fms
#superfamily macrophage colony-stimulating factor 1 receptor;
            immunoglobulin homology; protein kinase homology
KEYWORDS     ATP: autophosphorylation; glycoprotein; kinase-related
            transforming protein; magnesium; oncogene; phosphoprotein;
            phosphotransferase; receptor; transmembrane protein;
            tyrosine-specific protein kinase
FEATURE      #domain signal sequence #status predicted #label SIG\
1-23         #product protein-tyrosine kinase fms #status predicted
24-941       #label MAT\

24-509       #domain extracellular #status predicted #label EXT\
35-86        #domain immunoglobulin homology #label IMM1\
120-179      #domain immunoglobulin homology #label IMM2\
217-280      #domain immunoglobulin homology #label IMM3\
316-381      #domain immunoglobulin homology #label IMM4\
410-484      #domain immunoglobulin homology #label IMM5\
510-534      #domain transmembrane #status predicted #label TM\
535-941      #domain intracellular #status predicted #label INT\
577-915      #domain protein kinase homology #label KIN\
585-593      #region protein kinase ATP-binding motif\
42-84,127-177,
224-278,417-482,
457,73,94,153,275,
286,302,335,410,
477,490      #disulfide_bonds #status predicted\

613,630,776  #binding_site carbohydrate (Asn) (covalent) #status
781,794       predicted\
#active_site Lys, Glu, Asp #status predicted\
#binding_site magnesium (Asn, Asp) #status predicted\
SUMMARY      #length 941 #molecular_weight 104711 #checksum 3513

Query Match 58.1% Score 50; DB 1; Length 941;
Best Local Similarity 53.8%; Pred. No. 1.06e+01;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 852 NPYPGILVNSKFY 864
   : | | | | | |
QY 66 DPSTGALVDSKSY 78

RESULT 14
ENTRY   TVHUMD
TITLE   macrophage colony-stimulating factor 1 receptor precursor -
        human
CONTAINS protein-tyrosine kinase (EC 2.7.1.112) csflr/fms
ORGANISM #formal_name Homo sapiens #common_name man
DATE      28-Dec-1987 #sequence_revision 31-Dec-1993 #text_change
24-Oct-1997

ACCESSIONS S08123; A24533; I56672; I57648; I59083; I52772
REFERENCE   Hampe, A.; Shamoan, B.M.; Gobet, M.; Sherr, C.J.; Galibert,
#authors    F.
#journal     Oncogene Res. (1989) 4:9-17
#title       Nucleotide sequence and structural organization of the human

```

```

#cross-references MIM:89239490
#accession   S08123
#status      nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues    1-972 #label HAM
#cross-references GB:U63963; EMBL:X14720; NID:g1915975; PID:g1915976
#note        this sequence was submitted to the EMBL Data Library,
            March 1989

REFERENCE   A24533
#authors     Coussens, L.; Van Beveren, C.; Smith, D.; Chen, E.; Mitchell,
            R.L.; Isacke, C.M.; Verma, I.M.; Ullrich, A.
#journal     Nature (1986) 320:277-280
#title       Structural alteration of viral homologue of receptor
            proto-oncogene fms at carboxyl terminus.
#cross-references MIM:86175013
#accession   A24533
#molecule_type mRNA
#residues    1-53, 'A', '55-972 #label COU
#cross-references GB:J03149
#note        the authors translated the codon GCA for residue 54 as
            Pro

REFERENCE   I56672
#authors     Wheeler, E.F.; Rousset, M.F.; Hampe, A.; Walker, M.H.; Fried,
            V.A.; Look, A.T.; Rettenmier, C.W.; Sherr, C.J.
#journal     J. Virol. (1986) 59:224-233
#title       The amino-terminal domain of the v-fms oncogene product
            includes a functional signal peptide that directs synthesis
            of a transforming glycoprotein in the absence of feline
            leukemia virus gag sequences.
#cross-references MIM:86281820
#accession   I56672
#status      preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues    1-16 #label RES
#cross-references GB:M14002; NID:g182676; PID:g553292
REFERENCE   I57648
#authors     Vlasader, J.; Verma, I.M.
#journal     Mol. Cell. Biol. (1989) 9:1336-1341
#title       Differential transcription of exon 1 of the human c-fms gene
            in placental trophoblasts and monocytes.
#cross-references MIM:89261741
#accession   I57648
#status      preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues    1-16 #label RE2
#cross-references GB:M25786; NID:g349454; PID:g553224
REFERENCE   I59083
#authors     Browning, P.J.; Bunn, H.F.; Cline, A.; Shuman, M.; Nienhuis,
            A.W.
#journal     Proc. Natl. Acad. Sci. U.S.A. (1986) 83:7800-7804
#title       Replacement of COOH-terminal truncation of v-fms with c-fms
            sequences markedly reduces transformation potential.
#cross-references MIM:87017034
#accession   I59083
#status      translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues    874-972 #label RE3
#cross-references GB:M14193; NID:g182521; PID:g182522
REFERENCE   I52772
#authors     Nienhuis, A.W.; Bunn, H.F.; Turner, P.H.; Gopal, T.V.; Nash,
            W.G.; O'Brien, S.
#journal     Cell (1985) 42:421-428
#title       Expression of the human c-fms proto-oncogene in hematopoietic
            cells and its deletion in the 5q- syndrome.
#cross-references MIM:85282599
#accession   I52772
#status      preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues    244-295 #label RE4
#cross-references GB:M11067; NID:g182674; PID:g442423
GENETICS
#gene        GDB:CSF1R; FMS

```

```

##cross-references GDB:120600; OMIM:164770
#map_position 5q33.2-5q33.3
#introns
542/3; 585/1; 198/1; 243/3; 297/1; 361/2; 400/1; 440/2; 504/1;
542/3; 585/1; 620/1; 657/1; 711/2; 741/1; 773/3; 814/3;
852/1; 885/2; 921/3
CLASSIFICATION
#superfamily macrophage colony-stimulating factor 1 receptor;
immunoglobulin homology; protein kinase homology
KEYWORDS
ATP; autophosphorylation; glycoprotein; kinase-related
transforming protein; magnesium; phosphoprotein;
phosphotransferase; proto-oncogene; receptor; transmembrane
protein; tyrosine-specific protein kinase
FEATURE
1-23      #domain signal sequence #status predicted #label SIG\
24-972    #product macrophage colony-stimulating factor 1 receptor
          #status predicted #label MAT\
24-512    #domain extracellular #status predicted #label EXT\
35-86     #domain immunoglobulin homology #label IMM1\
120-179   #domain immunoglobulin homology #label IMM2\
217-280   #domain immunoglobulin homology #label IMM3\
316-383   #domain immunoglobulin homology #label IMM4\
412-487   #domain immunoglobulin homology #label IMM5\
513-537   #domain transmembrane #status predicted #label TM\
538-972   #domain intracellular #status predicted #label INT\
580-917   #domain protein kinase homology #label KIN\
588-596   #region protein kinase ATP-binding motif\
42-84,127-177,
224-278,419-485
45,73,153,240,275,
302,335,353,412,
428,480   #disulfide_bonds #status predicted\
          #binding_site carbohydrate (Asn) (covalent) #status
          predicted\
616,633,778 #active_site Lys, Glu, Asp #status predicted\
783,796     #binding_site magnesium (Asn, Asp) #status predicted
          #length 972 #molecular-weight 107983 #checksum 2888
SUMMARY
Query Match      58.1%; Score 50; DB 1; Length 972;
Best Local Similarity 53.8%; Pred. No. 1.06e+01;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 854 NPYPGILVNSKFY 866
   :| | | | |
Qy 66 DPSTGALVDSKSY 78

RESULT 15
ENTRY TVCTMD #type complete
TITLE macrophage colony-stimulating factor 1 receptor precursor -
cat
CONTAINS protein-tyrosine kinase (EC 2.7.1.112) csfir/fms
ORGANISM #formal name Felis silvestris catus #common_name domestic cat
DATE 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_Change
13-Jun-1997
ACCESSIONS A31636
REFERENCE Woolford, J.; McAuliffe, A.; Rohrschneider, L.R.
#authors Cell (1988) 55:965-977
#journal
#title Activation of the feline c-fms proto-oncogene: multiple
alterations are required to generate a fully transformed
phenotype.
#cross-references MUID:89077553
#accession A31636
#molecule_type mRNA
#residues 1-980 #label WOO
##cross-references EMBL:X03663
GENETICS
#gene fms
CLASSIFICATION #superfamily macrophage colony-stimulating factor 1 receptor;
immunoglobulin homology; protein kinase homology
KEYWORDS ATP; autophosphorylation; glycoprotein; kinase-related
transforming protein; magnesium; phosphoprotein;
phosphotransferase; proto-oncogene; receptor; transmembrane
protein; tyrosine-specific protein kinase
FEATURE

```

```

1-23      #domain signal sequence #status predicted #label SIG\
24-980    #product macrophage colony-stimulating factor 1 receptor
          #status predicted #label MAT\
24-509    #domain extracellular #status predicted #label EXT\
35-86     #domain immunoglobulin homology #label IMM1\
120-179   #domain immunoglobulin homology #label IMM2\
217-280   #domain immunoglobulin homology #label IMM3\
316-381   #domain immunoglobulin homology #label IMM4\
410-484   #domain immunoglobulin homology #label IMM5\
510-534   #domain transmembrane #status predicted #label TM\
535-980   #domain intracellular #status predicted #label INT\
577-915   #domain protein kinase homology #label KIN\
585-593   #region protein kinase ATP-binding motif\
42-84,127-177,
224-278,417-482
45,73,94,153,275,
302,335,410,477,
490       #disulfide_bonds #status predicted\
          #binding_site carbohydrate (Asn) (covalent) #status
          predicted\
613,630,776 #active_site Lys, Glu, Asp #status predicted\
781,794     #binding_site magnesium (Asn, Asp) #status predicted
          #length 980 #molecular-weight 108506 #checksum 7338
SUMMARY
Query Match      58.1%; Score 50; DB 1; Length 980;
Best Local Similarity 53.8%; Pred. No. 1.06e+01;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 852 NPYPGILVNSKFY 864
   :| | | | |
Qy 66 DPSTGALVDSKSY 78

Search completed: Wed Sep 1 16:30:13 1999
Job time : 11 secs.

```

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd
(TM)

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Sep 1 16:28:57 1999; MasPar time 2.70 seconds
Tabular output not generated. 146.747 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (66-79) from PCTUS9913024.pep (11 of 12)
Perfect Score: 86
Sequence: 1 DPSTGALVDSKSYA 14

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 25.367; Variance 25.400; scale 0.999

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	52	60.5	285	1	GLK2_XENLA GLUTAMATE RECEPTOR, IO	1.22e+00
2	52	60.5	899	1	GLK2_MOUSE GLUTAMATE RECEPTOR, IO	1.22e+00
3	52	60.5	908	1	GLK2_HUMAN GLUTAMATE RECEPTOR, IO	1.22e+00
4	52	60.5	908	1	GLK2_RAT GLUTAMATE RECEPTOR, IO	1.22e+00
5	52	60.5	918	1	GLK1_HUMAN GLUTAMATE RECEPTOR, IO	1.22e+00
6	52	60.5	919	1	GLK3_RAT GLUTAMATE RECEPTOR, IO	1.22e+00
7	52	60.5	919	1	GLK3_HUMAN GLUTAMATE RECEPTOR, IO	1.22e+00
8	52	60.5	949	1	GLK1_RAT GLUTAMATE RECEPTOR, IO	1.22e+00
9	51	59.3	469	1	CRTJ_RHOCA GLUTAMATE RECEPTOR, IO	1.22e+00
10	51	59.3	563	1	MUTL_VIBCH GLUTAMATE RECEPTOR, IO	1.22e+00
11	50	58.1	331	1	ASPO_ACIOL GLUTAMINASE-ASPARAGINA	2.10e+00
12	50	58.1	972	1	KFMS_HUMAN MACROPHAGE COLONY STIM	3.57e+00
13	50	58.1	978	1	KFMS_FSVMD MACROPHAGE COLONY STIM	3.57e+00
14	50	58.1	980	1	KFMS_FELCA MACROPHAGE COLONY STIM	3.57e+00
15	49	57.0	235	1	YIP2_YEAST HYPOTHETICAL 27.2 KD P	6.03e+00
16	49	57.0	577	1	VGLE_PVRVI GLYCOPROTEIN G1 PRECUR	6.03e+00
17	49	57.0	627	1	HEML_OPSTA 5-AMINOLEVULINIC ACID	6.03e+00
18	49	57.0	635	1	HEML_CHICK SERINE-REPEAT ANTIGEN	6.03e+00
19	49	57.0	989	1	SERA_PLAFC INSULIN RECEPTOR SUBST	6.03e+00
20	49	57.0	1321	1	IRS2_MOUSE PROTEIN SCK (FRAGMENT)	1.01e+01
21	48	55.8	428	1	SCK_HUMAN EXTRACELLULAR SERINE P	1.01e+01
22	48	55.8	448	1	PRTE_BACNO CRYPTIC PLASMAID PROTEI	1.67e+01
23	47	54.7	205	1	CPPC_NEIGO	1.67e+01

24 47 54.7 225 1 GTK1_RAT GLUTATHIONE S-TRANSFER 1.67e+01
25 47 54.7 434 1 YADA_YERPS INVASIN PRECURSOR (OUT 1.67e+01
26 47 54.7 450 1 DHE4_LACBI NADP-SPECIFIC GLUTAMAT 1.67e+01
27 47 54.7 457 1 DHE4_AGABI NADP-SPECIFIC GLUTAMAT 1.67e+01
28 47 54.7 473 1 VL2_HPV16 MINOR CAPSID PROTEIN L 1.67e+01
29 47 54.7 502 1 SYFB_YEAST PHENYLALANYL-CITRA SYNT 1.67e+01
30 47 54.7 522 1 GDS1_YEAST GDS1 PROTEIN. 1.67e+01
31 47 54.7 547 1 CITA_KLEPN SENSOR KINASE CITA (EC 1.67e+01
32 47 54.7 550 1 HEMA_IAME6 HEMAGGLUTININ PRECURSO 1.67e+01
33 47 54.7 583 1 T2F1_FLAOK TYPE IIS RESTRICTION E 1.67e+01
34 47 54.7 640 1 HEMI_HUMAN 5-AMINOLEVULINIC ACID 1.67e+01
35 47 54.7 778 1 ACON_SCHPO ACONITATE HYDRATASE, M 1.67e+01
36 47 54.7 883 1 APCE_CYAPA PHYCIBILISOME LINKER P 1.67e+01
37 47 54.7 976 1 KFMS_MOUSE MACROPHAGE COLONY STIM 1.67e+01
38 47 54.7 978 1 KFMS_RAT MACROPHAGE COLONY STIM 1.67e+01
39 47 54.7 1356 1 KAB7_YEAST PROBABLE SERINE/THREON 1.67e+01
40 47 54.7 2233 1 RRPL_P13H4 RNA POLYMERASE BETA SU 1.67e+01
41 46 53.5 82 1 NUMM_MOUSE NADH-UBIQUINONE OXIDOR 2.75e+01
42 46 53.5 271 1 YKUB_YEAST HYPOTHETICAL 31.2 KD P 2.75e+01
43 46 53.5 343 1 FIBP_ADE07 FIBER PROTEIN. 2.75e+01
44 46 53.5 469 1 GLNA_STRCO GLUTAMINE SYNTHETASE (2.75e+01
45 46 53.5 979 1 GLK5_MOUSE GLUTAMATE RECEPTOR, IO 2.75e+01

ALIGNMENTS

RESULT 1
ID GLK2_XENLA STANDARD; PRT; 285 AA.
AC Q91755;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 2 (GLUTAMATE RECEPTOR 6) (GLUR-6) (FRAGMENT).
GN GRIK2 OR GLUR6.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
CC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN:
RX MEDLINE; 96303825.
RA ISHIMARU H., KAMBOJ R., AMBROSINI A., HENLEY J.M., SOLOVIEV M.M.,
RA SUDAN H., ROSSIER J., ABUTIDZE K., RAMPERSAD V., USHERWOOD P.N.R.,
RA BATESON A.N., BARNARD E.A.;
RT "A unitary non-NMDA receptor short subunit from Xenopus: DNA cloning and expression".
RL RECEPT. CHANNELS 4:31-49(1996).
CC -!- FUNCTION: L-GLUTAMATE ACTS AS AN EXCITATORY NEUROTRANSMITTER AT MANY SYNAPSES IN THE CENTRAL NERVOUS SYSTEM. THE POSTSYNAPTIC ACTIONS OF GLU ARE MEDIATED BY A VARIETY OF RECEPTORS THAT ARE NAMED ACCORDING TO THEIR SELECTIVE AGONISTS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL: X94116; E214777; -
DR PFAM; PF00060; lig_chan; 1.
KW RECEPTOR; POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN;
KW TRANSMEMBRANE.
FT NON_TER 1 1
FT TRANSMEM 16 36 POTENTIAL.
FT TRANSMEM 197 217 POTENTIAL.
FT CARBOHYD 128 128 POTENTIAL.
SQ SEQUENCE 285 AA; 32041 MW; 5C7529B1 CRC32;


```

Query Match      60.5%; Score 52; DB 1; Length 285;
Best Local Similarity 60.0%; Pred. No. 1.22e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 133 GGLIDSKGYG 142
QY 70 GALVDSKSYA 79

RESULT 2
ID GLK2_MOUSE STANDARD; PRT; 889 AA.
AC P39087; Q60933;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 2 PRECURSOR (GLUTAMATE RECEPTOR
DE 6) (GLUR-6) (GLUTAMATE RECEPTOR BETA-2) (GLUR BETA-2) (FRAGMENT).
GN GRK2 OR GLUR6.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-BRAIN;
RX MEDLINE: 94083547.
RA GREGOR P., O'HARA B.F., YANG X., UHL G.R.;
RT "Expression and novel subunit isoforms of glutamate receptor genes
RT Glur5 and Glur6."
RL NEUROREPORT 4:1343-1346(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92356790.
RA MORITA T., SAKIMURA K., KUSHIYA E., YAMAZAKI M., MEGURO H.,
RA ARAKI K., ABE T., MORI K.J., MISHINA M.;
RT "Cloning and functional expression of a cDNA encoding the mouse beta
RT 2 subunit of the kainate-selective glutamate receptor channel."
RL BRAIN RES. MOL. BRAIN RES. 14:143-146(1992).
RN [3]
RP SEQUENCE OF 584-695 FROM N.A., AND RNA EDITING.
RC STRAIN-BALB/C;
RX MEDLINE: 96312506.
RA HERB A., HIGUCHI M., SPRENGEL R., SEEBURG P.H.;
RT "Q/R site editing in kainate receptor Glur5 and Glur6 pre-mRNAs
RT requires distant intronic sequences."
RL PROC. NATL. ACAD. SCI. U.S.A. 93:1875-1880(1996).
RN [4]
RP SEQUENCE OF 560-585 FROM N.A., AND RNA EDITING.
RC TISSUE-BRAIN;
RX MEDLINE: 93213505.
RA KOEHLER M., BURNASHEV N., SAKMANN B., SEEBURG P.H.;
RT "Determinants of Ca2+ permeability in both TM1 and TM2 of high
RT affinity kainate receptor channels: diversity by RNA editing."
RL NEURON 10:491-500(1993).
CC -!- FUNCTION: L-GLUTAMATE ACTS AS AN EXCITATORY NEUROTRANSMITTER AT
CC MANY SYNAPSES IN THE CENTRAL NERVOUS SYSTEM. THE POSTSYNAPTIC
CC ACTIONS OF GLU ARE MEDIATED BY A VARIETY OF RECEPTORS THAT ARE
CC NAMED ACCORDING TO THEIR SELECTIVE AGONISTS. MAY BE INVOLVED IN
CC THE TRANSMISSION OF LIGHT INFORMATION FROM THE RETINA TO THE
CC HYPOTHALAMUS.
CC -!- THE PRESENCE OF GLUTAMINE RESIDUE (NON-EDITED) DETERMINES CHANNELS
CC WITH LOW CALCIUM PERMEABILITY, WHEREAS AN ARGININE RESIDUE
CC (EDITED) DETERMINES A HIGHER CALCIUM PERMEABILITY ESPECIALLY IF
CC THE PRECEDING SITES ARE FULLY EDITED (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SUBUNIT: ASSEMBLES INTO A KAINATE-GATED HOMOMERIC CHANNEL THAT DOES
CC NOT BIND AMPA. GRK2 ASSOCIATED TO GRK5 FORMS FUNCTIONAL
CC CHANNELS. THAT CAN BE GATED BY AMPA (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: MOST ABUNDANT IN THE CEREBELLUM AND THE
CC HYPOTHALAMUS.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED DURING BRAIN DEVELOPMENT.
CC EXPRESSION DROPS IN THE ADULT.
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS GLUR6 BETA2 AND GLUR6-2 ARE
CC PRODUCED BY ALTERNATIVE SPLICING.

```

```

-!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; X66117; G312494; -
DR EMBL; D10054; D1001409; -
DR EMBL; U31443; G951154; -
DR PIR; S35792; S35792.
DR PIR; A43954; A43954.
DR MGI; 95815; GRK2.
DR PFAM; PF00600; lig_chan; 1.
DR RECEPTOR; POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL;
KW TRANSMEMBRANE; ALTERNATIVE SPLICING; RNA EDITING.
FT SIGNAL 1 31
FT CHAIN 32 >889
FT DOMAIN 32 561
FT TRANSMEM 562 582
FT TRANSMEM 601 621
FT TRANSMEM 636 656
FT TRANSMEM 820 840
FT CARBOHYD 67 67
FT CARBOHYD 73 73
FT CARBOHYD 275 275
FT CARBOHYD 378 378
FT CARBOHYD 412 412
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 546 546
FT VARSPIC 855 869
FT VARSPIC 870 >889
FT VARIANT 567 567
FT VARIANT 571 571
FT VARIANT 621 621
FT CONFLICT 611 611
FT CONFLICT 849 889
FT NON_TER 889
FT SEQUENCE 889 AA; 100273 MW; 32E5917A CRC32;

Query Match      60.5%; Score 52; DB 1; Length 889;
Best Local Similarity 60.0%; Pred. No. 1.22e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 756 GGLIDSKGYG 765
QY 70 GALVDSKSYA 79

RESULT 3
ID GLK2_HUMAN STANDARD; PRT; 908 AA.
AC Q13002;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 2 PRECURSOR (GLUTAMATE RECEPTOR
DE 6) (GLUR-6) (EXCITATORY AMINO ACID RECEPTOR 4) (EAA4).
GN GRK2 OR GLUR6.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-FETAL BRAIN;
RX MEDLINE: 95236039.
RA HOO K.H., NUTT S.L., FLETCHER E.J., ELLIOTT C.E., KORCZAK B.,
RA DEVERILL R.M., RAMPERSAD V., FANTASKE R.P., KAMBOJ R.K.;

```

RT "Functional expression and pharmacological characterization of the human EAA4 (GluR6) glutamate receptor: a kainate selective channel subunit.";
 RL RECEPT. CHANNELS 2:327-337(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94307730.
 RA PASCHEN W., BLACKSTONE C.D., HUGANIR R.L., ROSS C.A.;
 RT "Human GluR6 kainate receptor (GRIK2): molecular cloning, expression, polymorphism, and chromosomal assignment.";
 RL GENOMICS 20:435-440(1994).
 RN [3]
 RP RNA EDITING
 RX MEDLINE: 95016699.
 RA PASCHEN W., HEDREEN J.C., ROSS C.A.;
 RT "RNA editing of the glutamate receptor subunits GluR2 and GluR6 in human brain tissue.";
 RL J. NEUROCHEM. 63:1596-1602(1994).
 RN [4]
 RP RNA EDITING
 RC TISSUE-BRAIN;
 RX MEDLINE: 95210645.
 RA NUTT S.L., KAMBOJ R.K.;
 RT "RNA editing of human kainate receptor subunits.";
 RL NEUROREPORT 5:2625-2629(1994).
 CC -!- FUNCTION: L-GLUTAMATE ACTS AS AN EXCITATORY NEUROTRANSMITTER AT MANY SYNAPSES IN THE CENTRAL NERVOUS SYSTEM. THE POSTSYNAPTIC ACTIONS OF GLU ARE MEDIATED BY A VARIETY OF RECEPTORS THAT ARE NAMED ACCORDING TO THEIR SELECTIVE AGONISTS. MAY BE INVOLVED IN THE TRANSMISSION OF LIGHT INFORMATION FROM THE RETINA TO THE HYPOTHALAMUS. THIS RECEPTOR BINDS DOMOATE > KAINATE > QUISQUALATE > 6-CYANO-7-NITROQUINOLAXINE-2,3-DIONE > L-GLUTAMATE - 6,7-DINITROQUINOLAXINE-2,3-DIONE > DIHYDROKAINATE.
 CC -!- THE PRESENCE OF GLUTAMINE RESIDUE (NON-EDITED) DETERMINES CHANNELS WITH LOW CALCIUM PERMEABILITY, WHEREAS AN ARGININE RESIDUE (EDITED) DETERMINES A HIGHER CALCIUM PERMEABILITY ESPECIALLY IF THE PRECEDING SITES ARE FULLY EDITED. THIS RECEPTOR IS NEARLY COMPLETELY EDITED IN ALL GRAY MATTER STRUCTURES (90% OF THE RECEPTORS), WHEREAS MUCH LESS EDITING OCCURS IN THE WHITE MATTER (10% OF THE RECEPTORS).
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- SUBUNIT: ASSEMBLE INTO A KAINATE-GATED HOMOMERIC CHANNEL THAT DOES NOT BIND AMPA. GRIK2 ASSOCIATED TO GRIK5 FORMS FUNCTIONAL CHANNELS, THAT CAN BE GATED BY AMPA (BY SIMILARITY).
 CC -!- TISSUE SPECIFICITY: EXPRESSION IS HIGHER IN CEREBELLUM THAN IN CEREBRAL CORTEX.
 CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U16126; G790532;
 DR MIM: 138244;
 DR PFAM: PF00060; lig_chan; 1.
 KW RECEPTOR: POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL; TRANSMEMBRANE; RNA EDITING.
 FT SIGNAL 1 31 POTENTIAL.
 FT CHAIN 32 908 GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 2.
 FT DOMAIN 32 561 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 562 582 POTENTIAL.
 FT TRANSMEM 601 621 POTENTIAL.
 FT TRANSMEM 636 656 POTENTIAL.
 FT TRANSMEM 820 840 POTENTIAL.
 FT CARBOHYD 67 67 POTENTIAL.
 FT CARBOHYD 73 73 POTENTIAL.
 FT CARBOHYD 275 275 POTENTIAL.
 FT CARBOHYD 378 378 POTENTIAL.
 FT CARBOHYD 412 412 POTENTIAL.

FT CARBOHYD 423 423 POTENTIAL.
 FT CARBOHYD 430 430 POTENTIAL.
 FT CARBOHYD 546 546 POTENTIAL.
 FT VARIANT 567 567 I -> V (IN RNA EDITED VERSION).
 FT VARIANT 571 571 Y -> C (IN RNA EDITED VERSION).
 FT VARIANT 621 621 Q -> R (IN RNA EDITED VERSION).
 FT CONFLICT 789 789 G -> S (IN REF. 2).
 SQ SEQUENCE 908 AA; 102583 MW; D0058718 CRC32;
 Query Match 60.5%; Score 52; DB 1; Length 908;
 Best Local Similarity 60.08; Pred. No. 1.22e+00;
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 Db 756 GGLIDSKGYG 765
 QY 70 GALVDSKSYA 79
 :|:|:|:|:
 RESULT 4
 ID GLK2_RAT STANDARD; PRT; 908 AA.
 AC P42260;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 2 PRECURSOR (GLUTAMATE RECEPTOR 6) (GLUR-6).
 DE GN GRIK2 OR GLUR6.
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-BRAIN;
 RX MEDLINE: 91287799.
 RA EGEJERG J., BETTLER B., HERMANS-BORGMEYER I., HEINEMANN S.F.;
 RT "Cloning of a cDNA for a glutamate receptor subunit activated by kainate but not AMPA.";
 RL NATURE 351:745-748(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE: 93354697.
 RA LOMELI H., WISDEN W., KOEHLER M., KEINAENEN K., SOMMER B., SEEBURG P.H.;
 RT "High-affinity kainate and domoate receptors in rat brain.";
 RL FEBS LETT. 307:139-143(1992).
 RN [3]
 RP SEQUENCE OF 560-585 FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE: 93213505.
 RA KOEHLER M., BURNASHEV N., SAKMANN B., SEEBURG P.H.;
 RT "Determinants of Ca2+ permeability in both TM1 and TM2 of high affinity kainate receptor channels: diversity by RNA editing.";
 RL NEURON 10:491-500(1993).
 CC -!- FUNCTION: L-GLUTAMATE ACTS AS AN EXCITATORY NEUROTRANSMITTER AT MANY SYNAPSES IN THE CENTRAL NERVOUS SYSTEM. THE POSTSYNAPTIC ACTIONS OF GLU ARE MEDIATED BY A VARIETY OF RECEPTORS THAT ARE NAMED ACCORDING TO THEIR SELECTIVE AGONISTS. MAY BE INVOLVED IN THE TRANSMISSION OF LIGHT INFORMATION FROM THE RETINA TO THE HYPOTHALAMUS. THIS RECEPTOR BINDS KAINATE > QUISQUALATE > GLUTAMATE. IT DOES NOT BIND AMPA WITHOUT COEXPRESSION WITH GRIK5.
 CC -!- THE PRESENCE OF GLUTAMINE RESIDUE (NON-EDITED) DETERMINES CHANNELS WITH LOW CALCIUM PERMEABILITY, WHEREAS AN ARGININE RESIDUE (EDITED) DETERMINES A HIGHER CALCIUM PERMEABILITY ESPECIALLY IF THE PRECEDING SITES ARE FULLY EDITED. THIS RECEPTOR IS NEARLY COMPLETELY EDITED IN ALL GRAY MATTER STRUCTURES (90% OF THE RECEPTORS).
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- SUBUNIT: ASSEMBLE INTO A KAINATE-GATED HOMOMERIC CHANNEL THAT DOES NOT BIND AMPA. GRIK2 ASSOCIATED TO GRIK5 FORMS FUNCTIONAL CHANNELS, THAT CAN BE GATED BY AMPA.
 CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IS FOUND IN THE OLFACTORY LOBE, PIRIFORM CORTEX, DENTATE GYRUS, HIPPOCAMPUS, GRANULAR CELL

CC LAYER OF THE CEREBELLUM, AND IN CAUDATE-PUTAMEN.
 CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z11548; G56282; -;
 CC EMBL: Z11715; G56280; -;
 CC PFAM: PF00060; lig_chan; 1.
 CC RECEPTOR: POSTSYNAPTIC MEMBRANE: IONIC CHANNEL; GLYCOPROTEIN; SIGNAL;
 CC TRANSMEMBRANE: RNA EDITING.
 CC SIGNAL 1 31 POTENTIAL.
 CC CHAIN 32 908 GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 2.
 CC DOMAIN 32 561 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 562 582 POTENTIAL.
 CC TRANSMEM 601 621 POTENTIAL.
 CC TRANSMEM 636 656 POTENTIAL.
 CC TRANSMEM 820 840 POTENTIAL.
 CC CARBOHYD 67 67 POTENTIAL.
 CC CARBOHYD 73 73 POTENTIAL.
 CC CARBOHYD 275 275 POTENTIAL.
 CC CARBOHYD 378 378 POTENTIAL.
 CC CARBOHYD 412 412 POTENTIAL.
 CC CARBOHYD 423 423 POTENTIAL.
 CC CARBOHYD 430 430 POTENTIAL.
 CC CARBOHYD 546 546 POTENTIAL.
 CC VARIANT 567 567 I -> C (IN RNA EDITED VERSION).
 CC VARIANT 571 571 Y -> C (IN RNA EDITED VERSION).
 CC VARIANT 621 621 Q -> R (IN RNA EDITED VERSION).
 CC SEQUENCE 908 AA; 102470 MW; CA86C64C CRC32;
 CC
 CC Query Match 60.5%; Score 52; DB 1; Length 908;
 CC Best Local Similarity 60.0%; Pred. No. 1.22e+00;
 CC Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 CC
 Db 756 GGLIDSKGYG 765
 QY 70 GALVDSKSYA 79
 RESULT 5
 ID GLK1 HUMAN STANDARD; PRT; 918 AA.
 AC P39086; Q13001;
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 1 PRECURSOR (GLUTAMATE RECEPTOR
 DE 5) (GLUR-5) (EXCITATORY AMINO ACID RECEPTOR 3) (EAA3).
 GN GRIK1 OR GLUR5.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE-RETINA;
 RX MEDLINE: 94083547.
 RA GREGOR P., O'HARA B.F., YANG X., UHL G.R.;
 RA "Expression and novel subunit isoforms of glutamate receptor genes
 RT Glur5 and Glur6.";
 RL NEUROREPORT 4:1343-1346(1993).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE-FETAL BRAIN;
 RX MEDLINE: 96172461.
 RA KORCZAK B., NUTT S.L., FLETCHER E.J., HOO K.H., ELLIOTT C.E.,
 RA RAMPERSAD V., MCWHINIE E.A., KAMBOJ R.K.;
 RT "CDNA cloning and functional properties of human glutamate receptor
 RT EAA3 (Glur5) in homomeric and heteromeric configuration.";

RL RECEPT. CHANNELS 3:41-49(1995).
 RN [3]
 RP RNA EDITING.
 RC TISSUE-BRAIN;
 RX MEDLINE: 95210645.
 RA NUTT S.L., KAMBOJ R.K.;
 RA "RNA editing of human kainate receptor subunits.";
 RL NEUROREPORT 5:2625-2629(1994).
 CC -!- FUNCTION: L-GLUTAMATE ACTS AS AN EXCITATORY NEUROTRANSMITTER AT
 CC MANY SYNAPSES IN THE CENTRAL NERVOUS SYSTEM. THE POSTSYNAPTIC
 CC ACTIONS OF GLU ARE MEDIATED BY A VARIETY OF RECEPTORS THAT ARE
 CC NAMED ACCORDING TO THEIR SELECTIVE AGONISTS. MAY BE INVOLVED IN
 CC THE TRANSMISSION OF LIGHT INFORMATION FROM THE RETINA TO THE
 CC HYPOTHALAMUS. THIS RECEPTOR BINDS DOMOATE > KAINATE > L-GLUTAMATE
 CC - OUISOUALATE > CNQX - DNOX > AMPA > DIHYDROKAINATE > NMDA.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- SUBUNIT: THE UNEDITED VERSION (Q) ASSEMBLES INTO A FUNCTIONAL
 CC KAINATE-GATED HOMOMERIC CHANNEL, WHEREAS THE EDITED VERSION (R) IS
 CC UNABLE TO PRODUCE CHANNEL ACTIVITY WHEN EXPRESSED ALONE. BOTH
 CC EDITED AND UNEDITED VERSIONS CAN FORM FUNCTIONAL CHANNELS WITH
 CC GRIK4 AND GRIK5 (BY SIMILARITY).
 CC -!- TISSUE SPECIFICITY: MOST ABUNDANT IN THE CEREBELLUM AND THE
 CC SUPRACHIASMATIC NUCLEI (SCN) OF THE HYPOTHALAMUS.
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST TWO ISOFORMS (GLUR5-1D OR 1 AND
 CC EAA3A OR 2) ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L19058; G455448; -;
 CC EMBL: U16125; G790530; -;
 CC MIM: 138245; -;
 CC PFAM: PF00060; lig_chan; 1.
 CC RECEPTOR: POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL;
 CC TRANSMEMBRANE; PHOSPHORYLATION; RNA EDITING; ALTERNATIVE SPLICING.
 CC SIGNAL 1 30 POTENTIAL.
 CC CHAIN 31 918 GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 1.
 CC DOMAIN 31 576 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 577 597 POTENTIAL.
 CC TRANSMEM 616 636 POTENTIAL.
 CC TRANSMEM 634 674 POTENTIAL.
 CC TRANSMEM 835 855 POTENTIAL.
 CC CARBOHYD 68 68 POTENTIAL.
 CC CARBOHYD 74 74 POTENTIAL.
 CC CARBOHYD 276 276 POTENTIAL.
 CC CARBOHYD 379 379 POTENTIAL.
 CC CARBOHYD 428 428 POTENTIAL.
 CC CARBOHYD 439 439 POTENTIAL.
 CC CARBOHYD 446 446 POTENTIAL.
 CC CARBOHYD 561 561 POTENTIAL.
 CC MOD_RES 725 725 PHOSPHORYLATION (BY PKC) (POTENTIAL).
 CC MOD_RES 761 761 PHOSPHORYLATION (BY PKC) (POTENTIAL).
 CC VARIANT 636 636 Q -> R (IN RNA EDITED VERSION).
 CC VARSPLIC 402 416 MISSING (IN ISOFORM 2).
 CC VARSPLIC 870 918 KCSFTSILTCORRQKETA (IN ISOFORM 2).
 CC KCSFTSILTCORRQKETA (IN ISOFORM 2).
 CC KCSFTSILTCORRQKETA (IN ISOFORM 2).
 CC R -> G (IN REF. 2).
 CC CONFLICT 281 281 R -> G (IN REF. 2).
 CC SEQUENCE 918 AA; 103980 MW; 4A5251FB CRC32;
 CC
 CC Query Match 60.5%; Score 52; DB 1; Length 918;
 CC Best Local Similarity 60.0%; Pred. No. 1.22e+00;
 CC Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 CC
 Db 771 GGLIDSKGYG 780
 QY 70 GALVDSKSYA 79

CC GRIK5 TO FORM HETEROMERIC RECEPTORS (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U16127; G790534; -
 DR EMBL; U16128; G790536; -
 DR EMBL; S69349; G545834; -
 DR MIM; 138243; -
 DR PFAM; PF00060; lig_chan; 1.
 KW RECEPTOR; POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL;
 KW TRANSMEMBRANE; RNA EDITING.
 FT SIGNAL 1 31
 FT CHAIN 32 919
 FT DOMAIN 32 563
 FT TRANSMEM 564 584
 FT TRANSMEM 637 657
 FT TRANSMEM 821 841
 FT CARBOHYD 70 70
 FT CARBOHYD 76 76
 FT CARBOHYD 278 278
 FT CARBOHYD 381 381
 FT CARBOHYD 415 415
 FT CARBOHYD 426 426
 FT CARBOHYD 433 433
 FT CARBOHYD 548 548
 FT VARIANT 310 310
 FT VARIANT 352 352
 FT CONFLICT 303 303
 SQ SEQUENCE 919 AA: 104031 MW: 100CB78C0 CRC32:
 Query Match 60.5%; Score 52; DB 1; Length 919;
 Best Local Similarity 60.0%; Pred. No. 1.22e+00;
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 Db 757 GGLIDSKGYC 766
 |::|::|::|
 QY 70 GALVDSKSYA 79
 RESULT 8
 ID GLK1_RAT STANDARD; PRT; 949 AA.
 AC P22756;
 DT 01-AUG-1991 (REL. 19, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 1 PRECURSOR (GLUTAMATE RECEPTOR
 DE 5) (GLUR-5).
 GN GRIK1 OR GLURS.
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE-BRAIN;
 RX MEDLINE; 92224905.
 RA SOMMER B., BURNASHEV N., VERDOORN T.A., KEINANEN K., SAKMANN B.,
 RA SEEBURG P.H.;
 RT "A glutamate receptor channel with high affinity for domoate and
 RL kainate".
 RN EMBO J. 11:1651-1656(1992).
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE; 91026041.
 RA BETTLER B., BOULTER J., HERMANS-BORGMEYER I., O'SHEA-GREENFIELD A.,

RA DENNERIS E.S., MOLL C., BORGMEYER U., HOLLMANN M., HEINEMANN S.;
 RT "Cloning of a novel glutamate receptor subunit, GluR5: expression in
 RT the nervous system during development.";
 RL NEURON 5:583-595(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE; 92354697.
 RA LOMELI H., WISEN W., KOEHLER M., KEINANEN K., SOMMER B.,
 RA SEEBURG P.H.;
 RT "High-affinity kainate and domoate receptors in rat brain.";
 RL FEBS LETT. 307:139-143(1992).
 CC -!- FUNCTION: L-GLUTAMATE ACTS AS AN EXCITATORY NEUROTRANSMITTER AT
 CC MANY SYNAPSES IN THE CENTRAL NERVOUS SYSTEM. THE POSTSYNAPTIC
 CC ACTIONS OF GLU ARE MEDIATED BY A VARIETY OF RECEPTORS THAT ARE
 CC NAMED ACCORDING TO THEIR SELECTIVE AGONISTS. MAY BE INVOLVED IN
 CC THE TRANSMISSION OF LIGHT INFORMATION FROM THE RETINA TO THE
 CC HYPOTHALAMUS. THIS RECEPTOR BINDS DOMOATE > QUISQUALATE - L-
 CC GLUTAMATE > CNQX >> AMPA.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- SUBUNIT: THE UNEDITED VERSION OF (Q) ASSEMBLES INTO A FUNCTIONAL
 CC KAINATE-GATED HOMOMERIC CHANNEL, WHEREAS THE EDITED VERSION (R) IS
 CC UNABLE TO PRODUCE CHANNEL ACTIVITY WHEN EXPRESSED ALONE. BOTH
 CC EDITED AND UNEDITED VERSIONS CAN FORM FUNCTIONAL CHANNELS WITH
 CC GRIK4 AND GRIK5.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN SUBSETS OF NEURONS THROUGHOUT THE
 CC DEVELOPING AND ADULT CENTRAL AND PERIPHERAL NERVOUS SYSTEMS. IN
 CC THE CNS PRINCIPALLY IN THE MEDIAL AMYGDALOID NUCLEI, MEDIAL
 CC HABENULAE, PYRIFORM AND CINGULATE CORICES, AND PURKINJE CELL
 CC LAYER. ALSO HIGHLY EXPRESSED IN EMBRYONIC AND ADULT DORSAL ROOT
 CC GANGLIA.
 CC -!- ALTERNATIVE PRODUCTS: THERE ARE AT LEAST FOUR VARIANTS: GLURS-2,
 CC GLURS-2A, GLURS-2B/GLURS-1 AND GLURS-2C (SHOWN HERE); THAT ARISE
 CC BY ALTERNATIVE SPLICING.
 CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M83560; G204388; -
 DR EMBL; M83561; G204390; -
 DR EMBL; Z11712; G56274; -
 DR EMBL; Z11713; G56276; -
 DR EMBL; Z11714; G56278; -
 DR PIR; JH0312; JH0312.
 DR PIR; S19808; S19808.
 DR PIR; S19807; S19807.
 DR PFAM; PF00060; lig_chan; 1.
 KW RECEPTOR; POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL;
 KW TRANSMEMBRANE; PHOSPHORYLATION; ALTERNATIVE SPLICING; RNA EDITING.
 FT SIGNAL 1 30
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD_RES 761 761
 FT VARIANT 636 636
 FT CHAIN 31 949
 FT DOMAIN 31 576
 FT TRANSMEM 577 597
 FT TRANSMEM 616 636
 FT TRANSMEM 654 674
 FT TRANSMEM 835 855
 FT CARBOHYD 68 68
 FT CARBOHYD 74 74
 FT CARBOHYD 276 276
 FT CARBOHYD 379 379
 FT CARBOHYD 428 428
 FT CARBOHYD 439 439
 FT CARBOHYD 446 446
 FT CARBOHYD 561 561
 FT MOD_RES 725 725
 FT MOD

FT VARSPLIC 402 416 MISSING (IN GLURS-2).
FT VARSPLIC 870 898 MISSING (IN GLURS-2B/GLURS-1 AND
GLURS-2).
FT VARSPLIC 870 871 KG -> HY (IN GLURS-2A).
FT VARSPLIC 872 949 MISSING (IN GLURS-2A).
FT CONFLICT 282 282 K -> L (IN REF. 2).
FT CONFLICT 354 355 CA -> WR (IN REF. 2).
FT CONFLICT 477 477 A -> G (IN REF. 2).
SQ SEQUENCE 949 AA; 107840 MW; 7441125F CRC32;

Query Match 60.5%; Score 52; DB 1; Length 949;
Best Local Similarity 60.0%; Pred. No. 1.22e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 771 GGLIDSKYK 780
I::I::I::I::I::
Qy 70 GALVDSKSYA 79

RESULT 9
ID CRTJ-RHOCA STANDARD; PRT; 469 AA.
AC P26167;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE CRTJ PROTEIN.
GN CRTJ.
OS RHODOBACTER CAPSULATUS (RHODOPSEUDOMONAS CAPSULATA).
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHODOBACTER GROUP;
OC RHODOBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RA BURKE D.H., ALBERTI M., ARMSTRONG G.A., HEARST J.E.;
RL SUBMITTED (NOV-1991) TO EMBL/GENBANK/DDBJ DATA BANKS.
RC -!- PATHWAY: CAROTENOID AND CHLOROPHYLL BIOSYNTHESIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z11165; G995679; -.
DR
DR PIR; S17813; S17813.
DR PFAM; PF00158; sigma54; 1.
KW PHOTOSYNTHESIS; CHLOROPHYLL BIOSYNTHESIS; CAROTENOID BIOSYNTHESIS.
SQ SEQUENCE 469 AA; 51363 MW; 6C3D4D39 CRC32;

Query Match 59.3%; Score 51; DB 1; Length 469;
Best Local Similarity 42.9%; Pred. No. 2.10e+00;
Matches 6; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Db 308 DAGSAALVQGRSFA 321
I::I::I::I::I::
Qy 66 DPSTGALVDSKSYA 79

RESULT 10
ID MUTL-VIBCH STANDARD; PRT; 563 AA.
AC P11793;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PROTEIN MUTL.
GN MUTL.
OS VIBRIO CHOLERAE.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; VIBRIONACEAE; VIBRIO.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-569B.
RX MEDLINE; 89366654.

RA BERA T.K., GHOSH S.K., DAS J.;
RT "Cloning and characterization of mutL and mutS genes of Vibrio
RT cholerae: nucleotide sequence of the mutL gene.";
RL NUCLEIC ACIDS RES. 17:6241-6251(1989).
CC -!- FUNCTION: MUTL IS INVOLVED IN THE METHYL DIRECTED REPAIR OF
CC MISMATCHES IN DNA.
CC
CC -!- SIMILARITY: NONE WITH MUTL FROM OTHER BACTERIA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X15438; G48369; -.
DR PIR; S04886; S04886.
RW DNA REPAIR.
SQ SEQUENCE 563 AA; 64858 MW; 9F37C887 CRC32;

Query Match 59.3%; Score 51; DB 1; Length 563;
Best Local Similarity 53.6%; Pred. No. 2.10e+00;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 293 DPRTGVLVERK 303
I::I::I::I::I::
Qy 66 DPSTGALVDSK 76

RESULT 11
ID ASPO-ACIGL STANDARD; PRT; 331 AA.
AC P10172;
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE GLUTAMINASE-ASPARAGINASE (EC 3.5.1.38).
OS ACINETOBACTER GLUTAMINASIFICANS.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
OC MORAXELLACEAE; ACINETOBACTER.
RN [1]
RP SEQUENCE.
RX MEDLINE; 88243706.
RA TANAKA S., ROBINSON E.A., APPELLA E., MILLER M., AMMON H.L.,
RA ROBERTS J., WEBER I.T., WLODAWER A.;
RT "Structures of amidohydrolases. Amino acid sequence of a glutaminase-
RT asparaginase from Acinetobacter glutaminasificans and preliminary
RT crystallographic data for an asparaginase from Erwinia
RT chrysanthemi.";
RL J. BIOL. CHEM. 263:8583-8591(1988).
RN [2]
RP SEQUENCE OF 1-60.
RX MEDLINE; 78080774.
RA HOLCENBERG J.S., ERICSSON L., ROBERTS J.;
RT "Amino acid sequence of the diazooxonoleucine binding site of
RT Acinetobacter and Pseudomonas 7A glutaminase-asparaginase enzymes.";
RL BIOCHEMISTRY 17:411-417(1978).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
RA LUBKOWSKI J., WLODAWER A., HOUSSET D., WEBER I.T., AMMON H.L.,
RA MURPHY K.C., SWAIN A.L.;
RT "Refined crystal structure of Acinetobacter glutaminasificans
RT glutaminase-asparaginase";
RL ACTA CRYSTALLOGR. D 50:826-832(1994).
CC -!- CATALYTIC ACTIVITY: L-GLUTAMINE + H(2)O = L-GLUTAMATE + NH(3).
CC ALSO CATALYZES: L-ASPARAGINE + H(2)O = L-ASPARTATE + NH(3).
CC -!- SUBUNIT: HOMOTETRAMER.
DR PIR; A28063; A28063.
DR POB; 1AGX; 20-DEC-94.
DR PROSITE; PS00144; ASN_GLN_ASE_1; 1.
DR PROSITE; PS00917; ASN_GLN_ASE_2; 1.
DR PFAM; PF00710; Asparaginase; 1.
KW HYDROLASE; 3D-STRUCTURE.


```

QY      66 DPSTGALVDSKSY 78
RESULT  13
ID      KPM5.FSYMND      STANDARD;      PRT;      978 AA.
AC      P00545; Q86597;
DT      21-JUL-1986 (REL. 01, CREATED)
DT      01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT      15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE      FMS TYROSINE KINASE TRANSFORMING PROTEIN (EC 2.7.1.112).
GN      V-FMS.
OS      FELINE SARCOMA VIRUS (STRAIN MCDONOUGH).
OC      VIRUSES; RETROVIRUSES; RETROVIRIDAE; MAMMALIAN TYPE C RETROVIRUSES.
RN      [1]
RX      MEDLINE; 84119469.
RA      HAMPE A., GOBET M., SHERR C.J., GALIBERT F.;
RT      "Nucleotide sequence of the feline retroviral oncogene v-fms shows
RT      unexpected homology with oncogenes encoding tyrosine-specific protein
RT      kinases.";
RL      PROC. NATL. ACAD. SCI. U.S.A. 81:85-89(1984).
RN      [2]
RP      REVISIONS, SEQUENCE FROM N.A.
RX      MEDLINE; 92015516.
RA      SMOLA U., HENNIG D., HADWIGER-FANGMEIER A., SCHUETZ B., PFAFF E.,
RA      NIEMANN H., TAMURA T.;
RT      "Reassessment of the v-fms sequence: threonine phosphorylation of the
RT      COOH-terminal domain.";
RL      J. VIROL. 65:6181-6187(1991).
CC      -!- FUNCTION: V-FMS IS DERIVED FROM THE RECEPTOR FOR COLONY
CC      STIMULATING FACTOR 1 (CSF-1).
CC      -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC      -!- THIS PROTEIN IS SYNTHESIZED AS A GAG-FMS POLYPEPTIDE.
CC      -!- SIMILARITY: BELONGS TO THE CSF-1/PDGF RECEPTOR FAMILY OF TYROSINE-
CC      PROTEIN KINASES.
CC      -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. THE
CC      EXTRACELLULAR DOMAIN CONTAINS FIVE IG-FOLD DOMAINS.
CC
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL; K01643; G323891; -.
CC      DR      EMBL; S59588; G237024; -.
CC      DR      PIR; A00654; TMVMD.
CC      DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC      DR      PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC      DR      PROSITE; PS00240; RECEPTOR_TYR_KIN_III; 1.
CC      DR      PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC      DR      PFAM; PF00047; ig; 3.
CC      DR      PFAM; PF00069; pkinase; 2.
CC      DR      HSP; P11362; IFGI.
CC      KW      POLYPEPTIDE; TRANSFORMING PROTEIN; TYROSINE-PROTEIN KINASE; ONCOGENE;
CC      TRANSFERASE; RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; PHOSPHORYLATION;
CC      ATP-BINDING; IMMUNOGLOBULIN FOLD.
CC      FT      DOMAIN 1 543 EXTRACELLULAR (POTENTIAL).
CC      FT      TRANSMEM 544 588 POTENTIAL.
CC      FT      DOMAIN 569 978 CYTOPLASMIC (POTENTIAL).
CC      FT      DOMAIN 58 138 IG-LIKE DOMAIN 1.
CC      FT      DOMAIN 141 231 IG-LIKE DOMAIN 2.
CC      FT      DOMAIN 238 332 IG-LIKE DOMAIN 3.
CC      FT      DOMAIN 333 431 IG-LIKE DOMAIN 4.
CC      FT      DOMAIN 432 536 IG-LIKE DOMAIN 5.
CC      FT      DISULFID 76 118 POTENTIAL.
CC      FT      DISULFID 161 211 POTENTIAL.
CC      FT      DISULFID 258 312 POTENTIAL.
CC      FT      DISULFID 451 516 POTENTIAL.
CC      FT      DOMAIN 613 942 PROTEIN KINASE.
CC      FT      NP_BIND 619 627 ATP (BY SIMILARITY).

```

```

FT      BINDING      647      647      ATP (BY SIMILARITY).
FT      MOD_RES      841      841      PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT      MOD_RES      973      973      PHOSPHORYLATION.
FT      CARBOHYD      79      79      POTENTIAL.
FT      CARBOHYD      107      107      POTENTIAL.
FT      CARBOHYD      128      128      POTENTIAL.
FT      CARBOHYD      187      187      POTENTIAL.
FT      CARBOHYD      309      309      POTENTIAL.
FT      CARBOHYD      320      320      POTENTIAL.
FT      CARBOHYD      336      336      POTENTIAL.
FT      CARBOHYD      369      369      POTENTIAL.
FT      CARBOHYD      444      444      POTENTIAL.
FT      CARBOHYD      524      524      POTENTIAL.
FT      CONFLICT      714      714      L -> P (IN REF. 1).
FT      CONFLICT      971      978      ORTPVVAR -> RGPBL (IN REF. 1).
SQ      SEQUENCE      978 AA; 108491 MW; 2F1BCFB CRC32;

Query Match      58.1%; Score 50; DB 1; Length 978;
Best Local Similarity 53.8%; Pred. No. 3.57e+00;
Matches      7; Conservative      2; Mismatches      4; Indels      0; Gaps      0;

DB      886 NPYPGILVNSKFY 898
QY      66 DPSTGALVDSKSY 78
RESULT  14
ID      KPM5.FELCA      STANDARD;      PRT;      980 AA.
AC      F13369;
DT      01-JAN-1990 (REL. 13, CREATED)
DT      01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT      01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE      MACROPHAGE COLONY STIMULATING FACTOR I RECEPTOR PRECURSOR (CSF-1-R)
DE      (EC 2.7.1.112) (FMS PROTO-ONCOGENE) (C-FMS).
GN      CSF1R OR FMS.
OS      FELIS SILVESTRIS CATUS (CAT).
OC      EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC      CARNIVORA; FISSIPEDIA; FELIDAE; FELIS.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 89077553.
RA      WOOLFORD J., MCAULIFFE A., ROHRSCHEIDER L.R.;
RT      "Activation of the feline c-fms proto-oncogene: multiple alterations
RT      are required to generate a fully transformed phenotype.";
RL      CELL 55:965-977(1988).
CC      -!- FUNCTION: THIS PROTEIN IS THE RECEPTOR FOR CSF-1, IT IS A PROTEIN
CC      TYROSINE-KINASE TRANSMEMBRANE RECEPTOR.
CC      -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC      -!- SIMILARITY: BELONGS TO THE CSF-1/PDGF RECEPTOR FAMILY OF TYROSINE-
CC      PROTEIN KINASES.
CC      -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. THE
CC      EXTRACELLULAR DOMAIN CONTAINS FIVE IG-FOLD DOMAINS.
CC
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL; J03149; G163855; -.
CC      DR      PIR; A31636; TVTMD.
CC      DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC      DR      PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC      DR      PROSITE; PS00240; RECEPTOR_TYR_KIN_III; 1.
CC      DR      PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC      DR      PFAM; PF00047; ig; 3.
CC      DR      PFAM; PF00069; pkinase; 2.
CC      DR      HSP; P11362; IFGI.
CC      KW      TRANSFERASE; GLYCOPROTEIN; TYROSINE-PROTEIN KINASE; RECEPTOR; TRANSMEMBRANE;
CC      PROTO-ONCOGENE; TYROSINE-PROTEIN KINASE; PHOSPHORYLATION; ATP-BINDING; SIGNAL;

```


KW IMMUNOGLOBULIN FOLD. 19
FT SIGNAL 1
FT CHAIN 20 980
FT
FT DOMAIN 20 509
FT TRANSMEM 510 535
FT DOMAIN 536 980
FT DOMAIN 24 104
FT DOMAIN 107 197
FT DOMAIN 204 298
FT DOMAIN 299 397
FT DOMAIN 398 502
FT DOMAIN 579 908
FT NP_BIND 585 593
FT BINDING 613 613
FT ACT_SITE 776 776
FT DISULFID 42 84
FT DISULFID 127 177
FT DISULFID 224 278
FT DISULFID 417 482
FT MOD_RES 807 807
FT CARBOHYD 45 45
FT CARBOHYD 73 73
FT CARBOHYD 94 94
FT CARBOHYD 153 153
FT CARBOHYD 275 275
FT CARBOHYD 286 286
FT CARBOHYD 302 302
FT CARBOHYD 335 335
FT CARBOHYD 410 410
FT CARBOHYD 477 477
FT CARBOHYD 490 490
SQ SEQUENCE 980 AA; 108506 MW; EAICFFB6 CRC32;

Query Match 58.1%; Score 50; DB 1; Length 980;
Best Local Similarity 53.8%; Pred. No. 3.57e+00;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 852 NPYPGILVNSKFY 864
: | | | | |
QY 66 DPSTGALVDSKSY 78

RESULT 15
ID YIP2_YEAST STANDARD; PRT; 235 AA.
AC P40455;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 27.2 KD PROTEIN IN IMP2-DNA43 INTERGENIC REGION.
GN YIL152W.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
RA CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,
RA GENTLES S., HAMLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,
RA LOUIS E., LVE G., MOULE S., MOULE T., ODELL C., PEARSON D.,
RA RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,
RA WALSH S.V., WHITEHEAD S.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; 247047; G763194; -.
DR EMBL; 238059; G557771; -.
DR PIR; S48382; S48382.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 235 AA; 27201 MW; E0E5455D CRC32;

Query Match 57.0%; Score 49; DB 1; Length 235;
Best Local Similarity 46.2%; Pred. No. 6.03e+00;
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 207 DPSASLVDSRSF 219
| | | | | : | | | | |
QY 66 DPSTGALVDSKSY 78

Search completed: Wed Sep 1 16:29:07 1999
Job time : 10 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.

```

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:17:15 1999; MasPar time 6.29 Seconds
314.294 Million cell updates/sec

Tabular output not generated.

```

```

Title:
>PCT-US99-13024-2
Description:
(1-93) from PCTUS9913024.pep (6 of 12)
Perfect Score:
634
Sequence:
1 MEKFMAEFGGYVOTPFLE.....DKSYAFTSNDTTSAAFVS 93

Scoring table:
PAM 150
Gap 11

```

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

```
Database:
a-genseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39
```

Statistics: Mean 27.254; Variance 116.070; scale 0.235

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No
1	300	47.3	151	65	197317	Phage T4 ORFX gene pr	1.86e-18
2	95	15.1	1693	35	R73368	Hepatitis E virus hol	3.01e+00
3	92	14.5	520	1	P94617	Neutrial protease enco	6.20e+00
4	92	14.5	521	3	P51009	Sequence of neutral p	6.20e+00
5	87	13.7	424	19	P95000	Bacillus subtilis lpc	1.51e+01
6	87	13.7	623	35	P57773	Amino acid sequence o	1.51e+01
7	86	13.6	239	1	R05621	Hybrid (1.3-1,4)-pre-	1.80e+01
8	86	13.6	1693	34	W71209	Protein encoded by OR	1.80e+01
9	85	13.6	1693	19	R91813	Hepatitis E virus str	1.80e+01
10	85	13.6	1693	10	S12664	HEV strain protein en	1.80e+01
11	86	13.6	1693	37	W81519	Hepatitis E virus (HE	1.80e+01
12	85	13.6	1693	36	W80196	Protein encoded by OR	1.80e+01
13	85	13.4	633	35	P57774	Amino acid sequence o	2.14e+01
14	84	13.2	1693	3	R74618	Protein encoded by OR	2.55e+01
15	83	13.1	613	36	W73122	A. thaliana ethylene	3.04e+01
16	83	13.1	615	15	R74632	QETR ethylene respons	3.04e+01

```

AC W76358;
AD 03-DSC-1998 (first entry)
DE Hepatitis E virus hollow particle protein #1.
KW Hollow particle protein; virus; antibody; detection; immunoassay;
KW infection.
OS Hepatitis virus.
FH KeyWord(s): Location/Qualifiers
FT Protein
   1..1693
   /note= "Partial sequence"
FN J10234383-A.
PD 08-SEP-1998.
PF 28-FEB-1997; 0624445.
PR 28-FEB-1997; JP-062445.
PA (DENK-) DENKA SEIKEN KK.
PA (KOKU-) KOKURITSU YOBO EISEI KENKYUSHO.
DR WPI; 98-535037/46.
DR N-PSDB; V61687.
PT Hepatitis E virus hollow particle poly(peptide)s and nucleic acids
PT encoding it - useful for more accurate detection of HEV in samples,
PT using immuno-assays and nucleic acid hybridisation
PS Claim 10; Page 17-24; 29pp; Japanese.
CC This sequence represents a Hepatitis E viral hollow particle protein.
CC This polypeptides can be used to raise antibodies to detect HEV
CC Infection in samples, e.g. by immuno-assay based techniques, and the
CC nucleic acid can be used for the same in nucleic acid hybridisation
CC assays. The polypeptides and nucleic acids allow more accurate
CC detection of HEV than previously possible.
SQ Sequence 1693 AA;

Query Match      15.1% Score 96; DB 35; Length 1693;
Best Local Similarity 26.9%; Pred.No. 3.01e+00;
Matches 18; Conservative 21; Mismatches 25; Indels 3; Gaps

Db 572 frtsfdvgavleangperynlsfdaqsqtmaagpfstyaasaglevryvaagidhrav 631
QY 8 FGQGYVTPELSESNRYRKISIAAGSCLSTAGPSYVKFQDNPGVSGT-F-SAGLHLR-V 64
    |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 632 fapgvsp 638
QY 65 FDPSTGA 71
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 3
ID P94617 standard; protein; 520 AA.
AC P94617;
CD 21-JUN-1990 (first entry)
DE Neutral protease encoded by npr gene.
KE Protease; expression systems: subtilin; neutral protease; ds.
OS Bacillus amyloliquefaciens.
FH Key Word(s): Location/Qualifiers
FT protein
   221..520
   /note="Mature"
FT region
   27..520
   /note="PRO-"
FN USA801537-A.
PD 31-JAN-1989.
PF 29-MAR-1985; 717800.
PR 29-MAR-1985; US-717800.
PA (GENE-) Genex Corp.
PI Nagarajan V, Rhodes CS, Banner CDB;
DR WPI; 89-053639/07.
DR N-PSDB; N91114.
PT Vectors for expression of polypeptide(s) in Bacillus -
PT contg. promoter and regulatory regions which control expression
PT and secretion of protease(s) in Bacillus.
PS Disclosure; p: English.
CC Claimed replicon comprises a promoter and regulatory regions, capable
CC expressing alkaline and neutral protease genes.
SQ Sequence 520 AA;

Query Match      14.5% Score 92; DB 1; Length 520;
Best Local Similarity 27.8%; Pred.No. 6.20e+00;
Matches 15; Conservative 19; Mismatches 18; Indels 2; Gaps

```

```

Query Match      13.7%; Score 87; DB 19; Length 242;
Best Local Similarity 27.3%; Pred. No. 1.51e+01;
Matches 15; Conservative 19; Mismatches 18; Indels 3; Gaps 3;

Db 77 laitspsynkfcdgenrsvqtygygyerm-kpakntgivsfftytptdtp 130
QY 36 LSTAGPSYVKFQ-DNPVGSQTFSAGLH-LRVDFPSTGALVDSKSYAFSTSDNTS 88

RESULT
ID W75773 standard; Protein; 623 AA.
AC W75773.
DT 02-DEC-1998 (first entry)
DE Amino acid sequence of lepidoteran-active HD573 toxin.
KW HD573 toxin; PCR; primer; amplification; Bacillus thuringiensis; probe;
KW lepidoptera; pest; pesticide; Ostrinia nubilalis; Heliothis virescens;
KW Helicoverpa zea; hybridisation.
OS Bacillus thuringiensis.
PN W09840490-A1.
PD 17-SEP-1998.
PF 13-MAR-1998; U05081.
PR 13-MAR-1997; US-040512.
PA (MYCO ) MYCOGEN CORP.
PI Muller-Cohn J, Narva KE, Schnepf HE;
DR WPI; 98-506734/43.
DR N-PSDB; V52610.
PT New insecticidal Bacillus thuringiensis toxins - useful for
PT controlling lepidopteran pests, especially Ostrinia nubilalis,
PT Heliothis virescens and Helicoverpa zea
PS Claim 14; Pages 28-30; 50pp; English.
CC This is the amino acid sequence of a novel Bacillus thuringiensis toxin
CC used in the method of the invention, to control lepidopteran pests.
CC The new toxins are useful as pesticides, especially for the control of
CC Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The
CC polynucleotide coding sequences are useful for recombinant expression
CC of the toxins and the primers, together with probes derived from the
CC new sequences, are useful for the identification and characterisation
CC of novel genes that encode pesticidal toxins.
SQ Sequence 623 AA;

Query Match      13.7%; Score 87; DB 35; Length 623;
Best Local Similarity 18.2%; Pred. No. 1.51e+01;
Matches 12; Conservative 26; Mismatches 26; Indels 2; Gaps 2;

Db 502 fisekyngdgsrfelsnttarytlrgnsgsynlyrvssigsstirvingrytanv 561
QY 12 YVQTFPLSESNVRKYIS-IAGSCPLSTAGPSYVKF-QDNPVGSQTFSAGLHLRVDFPST 69

Db 562 nttnnn 567
QY 70 GALVDS 75

RESULT
ID R06621 standard; protein; 239 AA.
AC R06621;
DT 09-JAN-1991 (first entry)
DE Hybrid (1.3-1.4)-pre-beta-glucanase.
KW Hybrid pre-beta-glucanase; glucans; beer; animal feed; poultry.
OS Bacillus amyloliquefaciens, Bacillus macerans.
FH Key Location/Qualifiers
FT domain 1..129
FT /label-amino terminal of beta-amyloliquefaciens
FT domain 133..236
FT /label-carboxyl-terminal of B.macerans
PN W09009436-A.
PD 23-AUG-1990.
PF 16-FEB-1990; DK0044.
PR 16-FEB-1989; DD-325800.
PR 04-AUG-1989; DK-003848.
PA (CARL-) CARLSBERG A/S.
PA (DEAK ) AKAD WISSENSCHAFT DDR.

```

```

PI Borriass R, Hofemeister J, Thomsen KK, Olsen O, Vonwettstein D;
DR WPI; 90-275129/36.
DR N-PSDB; Q05832.
PT New thermostable (1.3-1.4)-beta-glucanase - prep'd. using hybrid
PT gene obt'd. using Bacillus amyloliquefaciens and B.macerans genes
PT disclosure; page 26; 84pp; English.
PS This hybrid protein is encoded by the beta-glucanase-H1 gene.
CC Following processing of the signal peptide the mature protein
CC is produced, comprising the amino terminus of the amylolique-
CC faciens beta-glucanase and the carboxyl-terminal half of the
CC B.macerans beta-glucanase. This hybrid protein is thermostable
CC and hydrolyses beta-glycosidic linkages in (1,3-1,4)-beta-glucans.
CC Reducing sugars are obt'd. at high temps. and thus this enzyme can
CC be used in the mfr. of food prods., esp. beer and animal feed (eg
CC for feeding poultry). See also Q05833.
SQ Sequence 239 AA;

Query Match      13.6%; Score 86; DB 1; Length 239;
Best Local Similarity 25.5%; Pred. No. 1.80e+01;
Matches 14; Conservative 20; Mismatches 18; Indels 3; Gaps 3;

Db 74 laitspsynkfcdgenrsvqtygygyerm-kpakntgivsfftytptdtp 127
QY 36 LSTAGPSYVKFQ-DNPVGSQTFSAGLH-LRVDFPSTGALVDSKSYAFSTSDNTS 88

RESULT
ID W71209 standard; Protein; 1693 AA.
AC W71209;
DT 30-OCT-1998 (first entry)
DE Protein encoded by ORF 1 of the Burmese isolate of ET-NANB.
KW Enterically transmitted nonA/nonB hepatitis virus; identification;
KW HEV; ET-NANB; detection; vaccine.
OS Hepatitis virus.
FH Key Location/Qualifiers
FT Misc_difference 154 /note= "not specified"
FT Misc_difference 1514 /note= "not specified"
FT Misc_difference 1552 /note= "not specified"
FT US789559-A.
PN 04-AUG-1998.
PD 25-JUN-1994; 279823.
PR 05-APR-1991; US-681078.
PR 17-JUN-1988; US-208997.
PR 11-APR-1989; US-336672.
PR 19-JUN-1989; US-367486.
PR 13-OCT-1989; US-420921.
PR 05-APR-1990; US-505888.
PR 25-JUL-1994; US-279823.
PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Bradley DW, Fry KE, Krawczynski KZ, Reyes GR, Tam A,
PI Yarbough PO;
DR WPI; 98-446186/38.
DR N-PSDB; V54729.
PT Hepatitis E virus DNA - useful for e.g. virus detection and viral
PT protein production
PS Disclosure; Columns 55-64; 45pp; English.
CC W71209-11 represent the proteins encoded by the open reading frames
CC (ORFs) of the DNA sequence of the Burmese isolate of an enterically
CC transmitted nonA/nonB viral hepatitis agent (ET-NANB). The nucleic
CC acid sequence may be used for identifying and sequencing the entire
CC viral agent (also referred to as HEV), detecting ET-NANB in
CC infected samples, e.g. by specific amplification of virus-derived DNA
CC sequences and for producing recombinant viral proteins for use in
CC vaccines.
SQ Sequence 1693 AA;

Query Match      13.6%; Score 86; DB 34; Length 1693;
Best Local Similarity 25.4%; Pred. No. 1.80e+01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

```

```
Db 572 frtsfdgavletngperhnlfsdasgstmaagpfsltayaasaglevryvaagldhrav 631
| :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: |
QY 8 FGQGYVOTPLSESNSVRYKISITAGSCLSTAGPSYVKFDNPVGSQT-F-SAGLHLR-V 64

Db 632 fapgvsp 638
| :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: |
QY 65 FDPSTGA 71

RESULT 9
ID R91813 standard; Protein; 1693 AA.
AC R91813;
DE Hepatitis E virus strain SAR-55 ORF-1.
DE Hepatitis E virus strain SAR-55 ORF-1.
DE Hepatitis E virus; HEV; SAR-55 strain; enteric transmission;
KW Structural region; antigen; detection; antibody; vaccine;
KW immunisation; infection.
OS Hepatitis E virus.
FH Key Location/Qualifiers
FT misc_difference 1238
FT misc_difference 1455..1693
FT /note= "corresponding codon CAG"
FT /note= "10 bp nucleic acid sequence TGGTNTTYGA
FT has to be inserted between nucleotides
FT 4390..4391 of T27394 before these amino
FT acid residues can be decoded"
FT
FT WO9610580-A2.
PD 11-APR-1996.
PF 03-OCT-1995; U13102.
PR 03-OCT-1994; US-316765.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Emerson SU, Purcell RH, Tsarev SA;
DR WPI: 96-209320/21.
DR N-PSDB: T27394.
PT Isolated and purified hepatitis E virus strain SAR-55 DNA - encodes
PT antigenic protein useful in diagnosis, prophylaxis and treatment of
PT hepatitis E virus infection
PS Disclosure; Pages 9-13; 121pp; English.
CC The present sequence is the protein prod. of ORF-1 from the
CC hepatitis E virus (HEV) strain SAR-55, which was implicated in an
CC enterically transmitted non-A, non-B hepatitis in Pakistan. The
CC protein encoded by the structural region of the virus (i.e. ORF-2),
CC which is capable of forming HEV like particles, is useful for the
CC detection of HEV antibodies (pref. IgG or IgM) in blood, plasma,
CC sera, cerebrospinal fluid, tissue, urine or pleural fluid. The
CC protein, and anti-HEV antibodies generated using the protein, can
CC also be used in vaccines for immunising an animal against HEV
CC infection. The protein is identified as a band of greater than
CC 50 kD following SDS-PAGE of cell lysates of insect cells infected
CC with a HEV ORF-2 contg. baculovirus, i.e. the claimed recombinant
CC expression vectors pPIC9-1779, -1780 and -1781.
CC Sequence 1693 AA;
SQ
Query Match 13.6%; Score 86; DB 19; Length 1693;
Best Local Similarity 25.4%; Pred. No. 1.80e+01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 frtsfdgavletngperhnlfsdasgstmaagpfsltayaasaglevryvaagldhrav 631
| :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: |
QY 8 FGQGYVOTPLSESNSVRYKISITAGSCLSTAGPSYVKFDNPVGSQT-F-SAGLHLR-V 64

Db 632 fapgvsp 638
| :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: |
QY 65 FDPSTGA 71

RESULT 10
ID R51264 standard; Protein; 1693 AA.
AC R51264;
DE Hepatitis E virus (first entry)
DE HEV strain protein encoded by ORF-1.
DE Hepatitis E virus; HEV; strain SAR-55; open reading frame; ORF;
KW antibody; detection; diagnosis; primates; stool suspension.
KW
```

```
OS Hepatitis E virus strain SAR-55.
PN WO9406913-A.
PD 31-MAR-1994.
PF 17-SEP-1993; U08849.
PR 18-SEP-1992; US-947263.
PA (USSH ) US SEC DEPT HEALTH.
PI Emerson SU, Purcell RH, Tsarev SA;
DR WPI: 94-118462/14.
DR N-PSDB: Q45197.
PT Purified hepatitis E strain SAR-55 virus - used to develop prods.
PT for use in detection, diagnosis, vaccines and therapy of
PT hepatitis E virus infection
PS Disclosure; Page 57-62; 114pp; English.
CC The sequences given in R51264-66 are encoded by the hepatitis E virus
CC (HEV) strain SAR-55. The cDNA sequence contains three open reading
CC frames (ORFs). These proteins can be used to stimulate the production
CC of protective antibodies upon injection into a mammal that would serve
CC to protect the mammal upon challenge with wild type HEV. The proteins
CC can be used for detection and diagnosis of HEV infection. The HEV
CC SAR-55 cDNA was isolated from primates inoculated with stool
CC suspensions obtained from hepatitis E patients.
CC Sequence 1693 AA;
SQ
Query Match 13.6%; Score 86; DB 10; Length 1693;
Best Local Similarity 25.4%; Pred. No. 1.80e+01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 frtsfdgavletngperhnlfsdasgstmaagpfsltayaasaglevryvaagldhrav 631
| :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: |
QY 8 FGQGYVOTPLSESNSVRYKISITAGSCLSTAGPSYVKFDNPVGSQT-F-SAGLHLR-V 64

Db 632 fapgvsp 638
| :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: | :|: |
QY 65 FDPSTGA 71

RESULT 11
ID W81519 standard; Protein; 1693 AA.
AC W81519;
DE Hepatitis E virus (first entry)
DE Hepatitis E virus (HEV) ORF-1 protein.
KW Hepatitis E virus; HEV; SAR-55; diagnostic agent; vaccine; antibody;
KW passive immunisation; open reading frame; ORF.
OS Hepatitis E virus.
FH Key Location/Qualifiers
FT Misc_difference 1238 /note= "encoded by CAG"
FT Misc_difference 1244 /note= "encoded by CAG"
FT Misc_difference 1352 /note= "encoded by CAG"
FT Misc_difference 1662 /note= "encoded by GGC"
FT Misc_difference 1662 /note= "encoded by GGC"
PN WO9846761-A1.
PD 22-OCT-1998.
PF 09-APR-1998; U07418.
PR 11-APR-1997; US-840316.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Emerson SU, Purcell RH, Robinson RA, Tsarev SA;
DR WPI: 98-568733/48.
DR N-PSDB: V71604.
PT New hepatitis E virus DNA from Pakistani strain SAR-55 - used for,
PT e.g. developing products for diagnosis of, and vaccination against
PT hepatitis E virus infection
PS Disclosure; Pages 11-15; 204pp; English.
CC This represents a hepatitis E virus (HEV) open reading frame (ORF)-1
CC protein encoded by a DNA sequence designated SAR-55. SAR-55 also encodes
CC HEV ORF-2 and ORF-3 proteins. A host organism transformed or transfected
CC with a recombinant expression vector containing the SAR-55 nucleic acid
CC can be used to produce the HEV proteins, especially ORF-2 protein. The
CC recombinant HEV proteins can be used as diagnostic agents and as vaccines
CC for use against HEV infection. The detection of antibodies specific for
CC HEV can be used for the diagnosis of infection and diseases caused by
```

CC HEV, and for monitoring the progression of such disease. Such methods are
 CC also useful for monitoring the efficacy of therapeutic agents during the
 CC course of treatment of HEV infection and disease in a mammal. The
 CC antibodies can be used for detection or for passive immunisation of
 CC mammals.
 SQ Sequence 1693 AA;

Query Match 13.6%; Score 86; DB 37; Length 1693;
 Best Local Similarity 25.4%; Pred. No. 1.80e+01;
 Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;
 Db 572 ftsfdvgavltngperhnlfsdastmaagpfslyaaasaaglevryvaagldhrav 631
 QY 8 FQGGVOTPFLESNSVRVKISAGSCLSTAGPSYVKFQDNVPGSQT-F-SAGLHLR-V 64
 Db 632 fapgvsp 638
 QY 65 FDPSTGA 71

RESULT 12
 ID W80196 standard; Protein: 1693 AA.
 AC W80196;
 DT 23-DEC-1998 (first entry)
 DE Protein encoded by ORF1 of ET-NANB (HEV) Burma strain DNA sequence.
 KW Enterically transmitted non A non B hepatitis virus; ET-NANB;
 KW Hepatitis E virus; HEV; Burma HEV isolate; vaccine;
 KW diagnostic probe.
 OS Non A non B Hepatitis virus.
 PN US5824649-A.
 PD 20-OCT-1998.
 PF 07-JUN-1995; 475807.
 PR 25-JUL-1994; US-279823.
 PR 17-JUN-1988; US-208997.
 PR 11-APR-1989; US-336672.
 PR 16-JUN-1989; US-367486.
 PR 13-OCT-1989; US-420921.
 PR 05-JUL-1990; US-505888.
 PR 07-JUN-1995; US-475807.
 PA (GENE-) GENELABS TECHNOLOGIES INC.
 PI Bradley DW, Fry KE, Krawczynski KZ, Reyes GR, Tam A,
 PI Yarbough PO;
 DR WPI: 98-582599/49.
 DR N-PSDB: V66321.

PT Hepatitis E virus proteins - useful for diagnosis or vaccine
 PS production of the virus
 PS Claim 22: Columns 57-66; 47pp; English.
 CC W80196-98 are encoded by the genome of the Burma strain of
 CC enterically transmitted non A non B hepatitis virus (ET-NANB)
 CC (hepatitis E virus (HEV)). The specification describes an isolated
 CC protein which is specifically immunoreactive with antibodies present
 CC in individuals infected with HEV and encoded by a sequence contained
 CC in an open reading frame (ORF) of an HEV genome. The genome has a
 CC sequence that is more than 70% identical to the ORF1 sequence from
 CC Burma HEV isolate. The protein is used as a vaccine and a diagnostic
 CC probe for ET-NANB.
 SQ Sequence 1693 AA;

Query Match 13.6%; Score 86; DB 36; Length 1693;
 Best Local Similarity 25.4%; Pred. No. 1.80e+01;
 Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 ftsfdvgavltngperhnlfsdastmaagpfslyaaasaaglevryvaagldhrav 631
 QY 8 FQGGVOTPFLESNSVRVKISAGSCLSTAGPSYVKFQDNVPGSQT-F-SAGLHLR-V 64
 Db 632 fapgvsp 638
 QY 65 FDPSTGA 71

RESULT 13
 ID W75774 standard; Protein: 633 AA.

AC W75774;
 DT 02-DEC-1998 (first entry)
 DE Amino acid sequence of lepidoteran-active HD525 toxin.
 KW HD525 toxin; PCR; primer; amplification; Bacillus thuringiensis; probe;
 KW lepidoptera; pest; pesticide; Ostrinia nubilalis; Heliothis virescens;
 KW Helicoverpa zea; hybridisation.
 OS Bacillus thuringiensis.
 PN WO9840490-A1.
 PD 17-SEP-1998.
 PF 13-MAR-1997; U05081.
 PR 13-MAR-1997; US-040512.
 PA (MYCO) MYCOGEN CORP.
 PI Muller-Cohn J, Narva KE, Schnepf HE;
 DR WPI: 98-506734/43.
 DR N-PSDB: V52611.
 DT New insecticidal Bacillus thuringiensis toxins - useful for
 PT controlling lepidopteran pests, especially Ostrinia nubilalis,
 PT Heliothis virescens and Helicoverpa zea
 PS Claim 14; Pages 32-34; 50pp; English.
 CC This is the amino acid sequence of a novel Bacillus thuringiensis toxin
 CC used in the method of the invention, to control lepidopteran pests.
 CC The new toxins are useful as pesticides, especially for the control of
 CC Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The
 CC polynucleotide coding sequences are useful for recombinant expression
 CC of the toxins and the primers, together with probes derived from the
 CC new sequences, are useful for the identification and characterisation
 CC of novel genes that encode pesticidal toxins.
 SQ Sequence 633 AA;

Query Match 13.4%; Score 85; DB 35; Length 633;
 Best Local Similarity 23.7%; Pred. No. 2.14e+01;
 Matches 14; Conservative 20; Mismatches 23; Indels 2; Gaps 2;

Db 511 fisekfngdsrlrfeqstntarytlrigngsynlyrvssignstirvtgrvytas 569
 QY 12 YVOTPFLESNSVRVKIS-IGSCPLSTAGPSYVKF-QDNVPGSQTFSAGLHLRVFDP 68

RESULT 14

ID R14618 standard; Protein: 1693 AA.
 AC R14618;
 DT 16-JAN-1992 (first entry)
 DE Protein encoded by ORF 1 of Burmese ET-NANB viral strain.
 KW Enterically transmitted non-A, non-B hepatitis virus; hepatitis C;
 KW HCV; E.coli strain BB4; ATCC deposit number 67717; Burma.
 OS Enterically transmitted non-A, non-B hepatitis virus.
 PN WO9115603-A.
 PD 17-OCT-1991.
 PF 05-APR-1991; U02368.
 PR 05-APR-1990; US-505888.
 PA (GENE-) GENELABS INC.
 PI (USSH) US DEPT HEALTH & HUMAN.
 PI Reyes GR, Yarbough PO, Bradley DW, Krawczynski KZ, Tam A;
 PI Fry KE;
 DR WPI: 91-325242/44.
 DR N-PSDB: Q14412.
 DT New viral proteins from non A-non-B hepatitis agent - used to
 PT treat and prevent enterically-transmitted non-A non-B hepatitis
 PT virus
 PS Disclosure; Page 15; 117pp; English.
 CC A positive clone ET1.1 was identified in a library prepared from
 CC bile of cynomolgus monkeys infected with the Burma strain of ET-NANB.
 CC Both strands of ET1.1 were sequenced. Identity of the sequence with
 CC sequences in etiologic agents has been confirmed by locating a
 CC similar sequence in a viral strain isolated in Burma. This protein
 CC is encoded by the longest ORF (ORF 1) of the Burma strain.
 CC (See Q14410 for ET1.1).
 SQ Sequence 1693 AA;

Query Match 13.2%; Score 84; DB 3; Length 1693;
 Best Local Similarity 26.6%; Pred. No. 2.55e+01;
 Matches 17; Conservative 20; Mismatches 24; Indels 3; Gaps 3;

Query Match 64.7%; Score 44; DB 2; Length 461;
Best Local Similarity 46.2%; Pred. No. 2.09e+01;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
Db 394 TTSADAVSSDFVG 406
QY 81 STSNDTTSAAVFS 93

RESULT 14
ID O44317 PRELIMINARY; PRT; 488 AA.
AC O44317;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE REVERSE TRANSCRIPTASE (FRAGMENT).
GN R1 ORF.
OS DUGESIELLA SP.
OC EUKARYOTA; METAZOA; ARTHROPODA; CHELICERATA; ARACHNIDA; ARANEAE;
RN [1]
RP SEQUENCE FROM N.A.
RA BURKE W.D., MALIK H.S., EICKBUSH T.H.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBBJ DATA BANKS.
DR EMBL; AF015813; G2736044; -;
KW RNA-DIRECTED DNA POLYMERASE.
FT NON_TER
SQ SEQUENCE 488 AA; 55092 MW; FA67B24C CRC32;

Query Match 64.7%; Score 44; DB 5; Length 488;
Best Local Similarity 58.3%; Pred. No. 2.09e+01;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 251 SKTSDGTGAFFV 262
QY 81 STSNDTTSAAVFS 92

RESULT 15
ID P87306 PRELIMINARY; PRT; 545 AA.
AC P87306;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 62.5 KD PROTEIN C31F10.04C IN CHROMOSOME II.
GN SPBC31F10.04C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA WOOD V., RAJANDREAM M.A., BARRELL B.G., POHL T.;
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBBJ DATA BANKS.
DR EMBL; Z97204; E325252; -;
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 545 AA; 62479 MW; 2F16D503 CRC32;

Query Match 64.7%; Score 44; DB 3; Length 545;
Best Local Similarity 54.5%; Pred. No. 2.09e+01;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 364 SNEDTSTNFVT 374
QY 83 SNDTTSAAVFS 93

Search completed: Wed Sep 1 16:32:18 1999
Job time : 34 secs.

RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RT eledans.;
 RL NATURE 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MILLER N.,
 RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA WATERSTON R.,
 RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: U00951; G107232;
 SQ SEQUENCE 53 AA; 5833 MW; CADA122E CRC32;
 Query Match 64.7%; Score 44; DB 5; Length 53;
 Best Local Similarity 58.3%; Pred. No. 2.09e+01;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 Db 6 STSNSTSRVSFV 17
 ||||| : :||
 QY 81 STSNDTTSAAVF 92
 RESULT 11
 ID Q00090 PRELIMINARY; PRT: 236 AA.
 AC Q00090
 DT 01-NOV-1996 (TREMREL. 01, CREATED)
 DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
 DE POTASSIUM CHANNEL PROTEIN (FRAGMENT).
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
 OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA RUDY B., KENTROS C., DE MIERA E.,
 RL MOL. CELL. NEUROSCI. 2:89-102(1991).
 CC -!- FUNCTION: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT POTASSIUM
 CC ION PERMEABILITY OF EXCITABLE MEMBRANES.
 CC ASSUMING OPENED OR CLOSED CONFORMATIONS IN RESPONSE TO THE
 CC VOLTAGE DIFFERENCE ACROSS THE MEMBRANE, THE PROTEIN FORMS A
 CC POTASSIUM-SELECTIVE CHANNEL THROUGH WHICH K+ IONS MAY PASS IN
 CC ACCORDANCE WITH THEIR ELECTROCHEMICAL GRADIENT.
 CC MAY PLAY A ROLE IN THE NERVOUS SYSTEM AND IN THE REGULATION OF
 CC BEATING FREQUENCY IN PACEMAKER CELLS.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- THE SEGMENT S4 IS PROBABLY THE VOLTAGE-SENSOR AND IS
 CC CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT
 CC EVERY THIRD POSITION.
 CC -!- THE AMINO TERMINUS MAY BE IMPORTANT IN DETERMINING THE RATE OF
 CC MODULATION OF THE CHANNEL WHILE THE TAIL MAY PLAY A ROLE IN
 CC INFLUENCING THE CHANNEL ACTIVITY AND/OR TARGETING OF THE CHANNEL TO
 CC SPECIFIC SUBCELLULAR COMPARTMENTS.
 CC -!- SIMILARITY: THIS CHANNEL PROTEIN BELONGS TO THE A-TYPE POTASSIUM
 CC CURRENT CLASS.
 CC -!- SIMILARITY: BELONGING TO THE SHIV GENE SUBFAMILY, A HOMOLOGUE OF
 CC THE SHAKER GENE SHAL OF DROSOPHILA MELANOGASTER.
 DR EMBL: M74898; G205041;
 DR PFAM: PF00520; Ion_trans; 1.
 KW IONIC CHANNEL; TRANSMEMBRANE; ION TRANSPORT; VOLTAGE-GATED CHANNEL;
 KW MULTIGENE FAMILY.
 FT NON_TER 1 1
 FT TRANSMEM 15 33 SEGMENT S1 (BY SIMILARITY).
 FT TRANSMEM 57 78 SEGMENT S2 (BY SIMILARITY).
 FT TRANSMEM 89 110 SEGMENT S3 (BY SIMILARITY).
 FT TRANSMEM 118 136 SEGMENT S4 (BY SIMILARITY).

FT TRANSMEM 152 173 SEGMENT S5 (BY SIMILARITY).
 FT TRANSMEM 213 234 SEGMENT S6 (BY SIMILARITY).
 FT NON_TER 236 236
 SQ SEQUENCE 236 AA; 26261 MW; C4A458FB CRC32;
 Query Match 64.7%; Score 44; DB 11; Length 236;
 Best Local Similarity 50.0%; Pred. No. 2.09e+01;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Db 108 TNNEDVSGAFVT 119
 ||||| : :|||
 QY 82 TSNDTTSAAFVS 93
 RESULT 12
 ID O65584 PRELIMINARY; PRT: 371 AA.
 AC O65584
 DT 01-AUG-1998 (TREMREL. 07, CREATED)
 DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 40.8 KD PROTEIN.
 GN M3E9.70.
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BEVAN M., VANDENBOL M., JALLET C., PORTETELLE D., HOHEISEL J.,
 RA MEWES H.W., MAYER K., SCHUELLER C.,
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU ARABIDOPSIS SEQUENCING PROJECT;
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AL022223; E1283943;
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 371 AA; 40833 MW; 580F5AA1 CRC32;
 Query Match 64.7%; Score 44; DB 10; Length 371;
 Best Local Similarity 38.5%; Pred. No. 2.09e+01;
 Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;
 Db 224 SSSGESSESSFSVS 236
 ||||| : :|||
 QY 81 STSNDTTSAAFVS 93
 RESULT 13
 ID O68923 PRELIMINARY; PRT: 461 AA.
 AC O68923
 DT 01-AUG-1998 (TREMREL. 07, CREATED)
 DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
 DE GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE HOMOLOG.
 GN GAPX.
 OS STREPTOMYCES ROSEOFULVUS.
 OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
 OC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94237486.
 RA BIBB M.J., SHERMAN D.H., OMURA S., HOPWOOD D.A.,
 RT "Cloning, sequencing and deduced functions of a cluster of
 RT Streptomyces genes probably encoding biosynthesis of the polyketide
 RT antibiotic frenolicin.";
 RL GENE 142:31-39(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA REEVES C.D., SOLIDAY C.L.,
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF058302; G3170587;
 SQ SEQUENCE 461 AA; 49966 MW; 26EC2672 CRC32;

```
OC EUKARYOTA: DICTYOSTELIIDA: DICTYOSTELIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AX4:
RA LOOMIS W.F.:
RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBDJ DATA BANKS.
DR EMBL: U66911; G1519334; -.
RN NON_TER 268 268
SQ SEQUENCE 268 AA; 30503 MW; 04E723F5 CRC32;

Query Match 67.6%; Score 46; DB 5; Length 268;
Best Local Similarity 53.8%; Pred. No. 7.24e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 15 TTNNNTTNNKFSV 27
      :|:|:|:|
Qy 81 STSNDTTSAAFFV 93

RESULT 7
ID Q19474 PRELIMINARY; PRT; 337 AA.
AC Q19474:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE F15A2.4 PROTEIN.
GN F15A2.4.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA: METAZOA: NEMATODA: SECERNENTEA: RHABDITIA: RHABDITIDA:
OC RHABDITINA: RHABDITOIDEA: RHABDITIIDAE: PELODERINAE: CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA GREGORY J.:
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBDJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KESHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.; III of C.
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL NATURE 368:32-38(1994).
DR EMBL: Z70207; E1345607; -.
SQ SEQUENCE 337 AA; 38483 MW; ABD004B8 CRC32;

Query Match 66.2%; Score 45; DB 5; Length 337;
Best Local Similarity 66.7%; Pred. No. 1.24e+01;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 317 SKHNSSTTSATFV 328
      | | | | | | | |
Qy 81 STSNDTTSAAFFV 92

RESULT 8
ID Q12372 PRELIMINARY; PRT; 583 AA.
AC Q12372:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CHROMOSOME XII READING FRAME ORF YLLO61W.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA: FUNGI: ASCOMYCOTA: HEMIASCOCYCETES: SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE: SACCHAROMYCES.
RN [1]

RP SEQUENCE FROM N.A.
RA WEDLER H., WEDLER E., SCHARFE M., WAMBUIT R.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBDJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA MIPS;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBDJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN-S288C;
RA WEDLER H., WAMBUIT R.;
RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBDJ DATA BANKS.
DR EMBL: Z73166; E245763; -.
DR EMBL: Z47973; G642323; -.
DR PFAM: PF00324; aa_permeases; 1.
SQ SEQUENCE 583 AA; 64217 MW; BDE5153C CRC32;

Query Match 66.2%; Score 45; DB 3; Length 583;
Best Local Similarity 50.0%; Pred. No. 1.24e+01;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 330 SSEDITASPFV 341
      |:|:| |::|
Qy 81 STSNDTTSAAFFV 92

RESULT 9
ID Q26818 PRELIMINARY; PRT; 37 AA.
AC Q26818:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SURFACE ANTIGEN (FRAGMENT).
OS TRYPAOSOMA BRUCEI BRUCEI.
OC EUKARYOTA: EULENZOZA: KINETOPLASTIDA: TRYPAOSOMATIDAE: TRYPAOSOMA.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-GAMBIENSE;
RX MERRITT S.C., TSCHUDI C., KONIGSBERG W.H., RICHARDS F.F.;
RT "Reverse transcription of trypanosome variable antigen mRNAs
RT initiated by a specific oligonucleotide primer."
RL PROC. NATL. ACAD. SCI. U.S.A. 80:1536-1540(1983).
DR EMBL: V01391; G10560; -.
FT NON_TER 1
SQ SEQUENCE 37 AA; 4068 MW; 551442D4 CRC32;

Query Match 64.7%; Score 44; DB 5; Length 37;
Best Local Similarity 63.6%; Pred. No. 2.09e+01;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 14 TSNTTASNSFV 24
      | | | | | | |
Qy 82 TSNDTTSAAFFV 92

RESULT 10
ID Q23582 PRELIMINARY; PRT; 53 AA.
AC Q23582:
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SIMILAR TO ADENOSINE A3 RECEPTOR. NCBI GI: 1072232.
GN ZK721.5.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA: METAZOA: NEMATODA: SECERNENTEA: RHABDITIA: RHABDITIDA:
OC RHABDITINA: RHABDITOIDEA: RHABDITIIDAE: PELODERINAE: CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
```

```
RP SEQUENCE FROM N.A.
RA POWERS P.A., GREGG R.G., HOGAN K.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U86953; G2155256; JOINED.
DR EMBL; U86954; G2155256; JOINED.
DR EMBL; U86955; G2155256; JOINED.
DR EMBL; U86956; G2155256; JOINED.
DR EMBL; U86957; G2155256; JOINED.
DR EMBL; U86958; G2155256; JOINED.
DR PFAM; PF00018; SH3; 1.
SQ SEQUENCE 579 AA; 63705 MW; 651614EF CRC32;

Query Match 79.4%; Score 54; DB 4; Length 579;
Best Local Similarity 75.0%; Pred. No. 7.52e-02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTSNSFV 55
   |||:||||:|
QY 81 STSDTTSAAVF 92

RESULT 3
ID Q20870 PRELIMINARY; PRT; 241 AA.
AC Q20870;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SIMILAR TO GLUTAMATE RECEPTORS.
GN F56D1.7.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDIIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPEX T., COOPER J., FULTON L.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., JIER M., JOHNSTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA CHISSOE S., WILSON R.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U39997; G1055122; -.
SQ SEQUENCE 241 AA; 27445 MW; F0977197 CRC32;

Query Match 69.1%; Score 47; DB 5; Length 241;
Best Local Similarity 80.0%; Pred. No. 4.20e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 59 GNDTISAAPV 68
   :||| |||||
QY 83 SNTTSAAPV 92

RESULT 4
ID Q09532 PRELIMINARY; PRT; 982 AA.
AC Q09532;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 110.9 KD PROTEIN F07F6.6 IN CHROMOSOME II.
GN F07F6.6.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDIIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA CHISSOE S.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: TO GLUTAMATE RECEPTOR.
DR EMBL; U23486; G746452; -.
DR WORMPEP; F07F6.6; CE01898.
DR PFAM; PF00060; lig_chan; 1.
KW HYPOTHETICAL PROTEIN.
FT DOMAIN 827 830 POLY-ARG.
SQ SEQUENCE 982 AA; 110892 MW; B4285770 CRC32;

Query Match 69.1%; Score 47; DB 5; Length 982;
Best Local Similarity 80.0%; Pred. No. 4.20e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 59 GNDTISAAPV 68
   :||| |||||
QY 83 SNTTSAAPV 92

RESULT 5
ID P87112 PRELIMINARY; PRT; 2100 AA.
AC P87112;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE PUTATIVE TRANSCRIPTIONAL REGULATOR C20G8.06.
GN SPAC20G8.06.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA BADCOCK K., CHURCHER C.M., WOOD V., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: TO YEAST NOT1.
DR EMBL; Z95334; E315491; -.
KW HYPOTHETICAL PROTEIN; NUCLEAR PROTEIN; TRANSCRIPTION REGULATION.
SQ SEQUENCE 2100 AA; 237132 MW; 9A710DB4 CRC32;

Query Match 69.1%; Score 47; DB 3; Length 2100;
Best Local Similarity 66.7%; Pred. No. 4.20e+00;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1601 SATNDETAAVF 1612
   :||| |||||
QY 81 STSNDTTSAAVF 92

RESULT 6
ID Q94484 PRELIMINARY; PRT; 268 AA.
AC Q94484;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)
DE ORF DG1007 (FRAGMENT).
OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
```

WATERMAN
***** (TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:31:44 1999; MasPar time 5.61 Seconds

Tabular output not generated. 126.558 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (81-93) from PCTUS9913024.pep (12 of 12)
Perfect Score: 68

Sequence: 1 STSNDTTSAAAFVS 13

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 21.936; Variance 22.856; scale 0.960

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	54	79.4	101	11	088517 CALCIUM CHANNEL BETA 1	7.52e-02
2	54	79.4	579	4	015331 L-TYPE CALCIUM CHANNEL	7.52e-02
3	47	69.1	241	5	Q20870 SIMILAR TO GLUTAMATE R	4.20e+00
4	47	69.1	982	5	Q09532 HYPOTHETICAL 110.9 KD	4.20e+00
5	47	69.1	2100	3	P87112 PUTATIVE TRANSCRIPTION	4.20e+00
6	46	67.6	268	5	Q94484 ORF DG1007 (FRAGMENT).	7.24e+00
7	45	66.2	337	5	Q19474 FL5A2.4 PROTEIN.	1.24e+01
8	45	66.2	583	3	Q12372 CHROMOSOME XII READING	1.24e+01
9	44	64.7	37	5	Q26818 SURFACE ANTIGEN (FRAGM	2.09e+01
10	44	64.7	53	5	Q23582 SIMILAR TO ADENOSINE A	2.09e+01
11	44	64.7	236	11	Q00090 POTASSIUM CHANNEL PROT	2.09e+01
12	44	64.7	371	10	Q65584 HYPOTHETICAL 40.8 KD P	2.09e+01
13	44	64.7	461	2	Q68923 GLYCERALDEHYDE-3-PHOSP	2.09e+01
14	44	64.7	488	5	Q44317 REVERSE TRANSCRIPTASE	2.09e+01
15	44	64.7	545	3	P87306 HYPOTHETICAL 62.5 KD P	2.09e+01
16	44	64.7	575	5	Q26040 SHAL 1 POTASSIUM CHANN	2.09e+01
17	44	64.7	611	11	Q63286 VOLTAGE-GATED K+ CHANN	2.09e+01
18	44	64.7	623	5	Q16992 W03F9.4 PROTEIN.	2.09e+01
19	44	64.7	636	11	P70622 KV4.3 POTASSIUM CHANNE	2.09e+01
20	44	64.7	636	4	Q60576 KV4.3 POTASSIUM CHANNE	2.09e+01

21	44	64.7	636	11	Q62897 SHAL-RELATED POTASSIUM	2.09e+01
22	44	64.7	655	4	Q60577 KV4.3 POTASSIUM CHANNE	2.09e+01
23	44	64.7	655	11	Q08723 KV4.3	2.09e+01
24	44	64.7	669	2	Q68385 FLAGELLIN (FRAGMENT).	2.09e+01
25	44	64.7	1322	5	Q18286 SIMILAR TO GTP-BINDING	2.09e+01
26	44	64.7	1419	3	Q13368 AGGLUTININ-LIKE ADHESI	2.09e+01
27	44	64.7	3110	4	Q93022 LAMININ ALPHA 2 CHAIN.	2.09e+01
28	43	63.2	110	3	Q12338 CHROMOSOME XII READING	3.51e+01
29	43	63.2	290	2	P96610 YDPO PROTEIN.	3.51e+01
30	43	63.2	316	5	Q01681 CODED FOR BY C. ELEGAN	3.51e+01
31	43	63.2	457	5	Q09579 HYPOTHETICAL 52.3 KD P	3.51e+01
32	43	63.2	488	10	P93694 PHEROPHORIN III (FRAGM	3.51e+01
33	43	63.2	490	11	Q99249 POTASSIUM CHANNEL PROT	3.51e+01
34	43	63.2	521	2	Q87767 GP5-LIKE PROTEIN (FRAG	3.51e+01
35	43	63.2	543	5	Q16749 SIMILARITY TO A HOMEOB	3.51e+01
36	43	63.2	549	2	Q34801 DPPE.	3.51e+01
37	43	63.2	552	2	Q25003 SODIUM-DEPENDENT TRANS	3.51e+01
38	43	63.2	630	11	Q13881 SHAL1.	3.51e+01
39	43	63.2	649	14	Q10619 IMMEDIATE EARLY PROTEI	3.51e+01
40	43	63.2	979	3	Q60079 UBQUITIN CARBOXYL-TER	3.51e+01
41	43	63.2	1229	5	Q94185 CODED FOR BY C. ELEGAN	3.51e+01
42	43	63.2	1231	4	Q75646 DJ232K4.1 (HYPOTHETICA	3.51e+01
43	43	63.2	1266	4	Q92833 HYPOTHETICAL 141.7 KD	3.51e+01
44	43	63.2	2450	11	Q62135 PROTEIN-TYROSINE PHOSP	3.51e+01
45	42	61.8	1103	14	Q41928 SSDNA BINDING PROTEIN.	5.83e+01

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	101 AA.
1	088517			
AC	088517:			
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	DE CALCIUM CHANNEL BETA 1 SUBUNIT (FRAGMENT).			
GN	CACNB1.			
OS	MUS MUSCULUS (MOUSE).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;			
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-NMRI; TISSUE-BRAIN;			
RA	HILDENBRAND J., AMMON H.P.T., WAHL M.A.;			
RT	"The mouse voltage-gated calcium channel beta 1 subunit, partial cDNA			
RT	sequence.";			
RL	SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.			
DR	EMBL; AF068898; G3202008; -			
FT	NON_TER 1			
FT	NON_TER 101			
SQ	SEQUENCE 101 AA; 11240 MW; 838619F3 CRC32;			

Query Match 79.4%; Score 54; DB 11; Length 101;
Best Local Similarity 75.0%; Pred. No. 7.52e-02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db	36	STSDTTSNSFV 47
Qy	81	STSDTTSAAFV 92

RESULT	ID	PRELIMINARY:	PRT:	579 AA.
2	015331			
AC	015331:			
DT	01-JAN-1998 (TREMBLREL. 05, CREATED)			
DT	01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	DE L-TYPE CALCIUM CHANNEL BETA SUBUNIT ISOFORM BETA-1B.			
GN	CACNB1			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;			
OC	CATARRHINI; HOMINIDAE; HOMO.			
RN	[1]			

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67505; G1591267;
DR TIGR; MJ0561;
DR PROSITE; PS01266; ADENYLOSUCCIN_SYN_1; 1.
DR PFAM; PF00709; ADENYLOSUCCIN_SYN_2; FALSE_NEG.
DR HSSP; P12283; IADI.
KW PURINE BIOSYNTHESIS; LIGASE; GTP-BINDING.
FT NP_BIND 18 24
SQ SEQUENCE 345 AA; 37820 MW; E15BEZC9 CRC32;

Query Match 66.2%; Score 45; DB 1; Length 345;
Best Local Similarity 50.0%; Pred. No. 5.97e+00;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 200 TSKDNTASSPRA 211
QY 82 TSNDTTSAAFS 93
||:||||:|

RESULT 14
ID IR12_HCMVA STANDARD; PRT; 416 AA.
AC P16810;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DE 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN IRL12 (TFL12).
OS HUMAN CYTOMEGALOVIRUS (STRAIN AD169).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC BETAHERPESVIRINAE; CYTOMEGALOVIRUS.
FN [1]
SEQUENCE FROM N.A.
RX MEDLINE; 90269039.
RA CHEE M.S., BANKIER A.T., BECK S., BOHNI R., BROWN C.M., CERNY R.,
RA HORSNELL T., HUTCHISON C.A. III, KOUZARIDES T., MARTIGNETTI J.A.,
RA PREDDIE E., SATCHWELL S.C., TOMLINSON P., WESTON K.M., BARRELL B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169."
RL CURR. TOP. MICROBIOL. IMMUNOL. 154:125-169(1990).
CC -1- SIMILARITY: BELONGS TO RL11 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17403; E298604;
DR EMBL; X17403; E298631;
DR PIR; S09761; S09761.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 416 AA; 47416 MW; 8289EB89 CRC32;

Query Match 66.2%; Score 45; DB 1; Length 416;
Best Local Similarity 50.0%; Pred. No. 5.97e+00;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 42 TSPNTASTFTVT 53
QY 82 TSNDTTSAAFS 93
||:||||:|

RESULT 15
ID POLG_BCMVN STANDARD; PRT; 3066 AA.
AC Q85399;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN [CONTAINS: N-TERMINAL PROTEIN (P1); HELPER
DE COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KD PROTEIN

```

```

DE 1 (6K1); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KD PROTEIN 2 (6K2);
DE GENOME-LINKED PROTEIN (VPC); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA)
DE (EC 3.4.22.-) (49 KD PROTEINASE) (49 KD-PRO); NUCLEAR INCLUSION
DE PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);
DE COAT PROTEIN (CP)].
OS BEAN COMMON MOSAIC VIRUS (STRAIN NL-3 / MICHIGAN) (BCMV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
OC POTYVIRUS.
FN [1]
SEQUENCE FROM N.A.
RX MEDLINE; 96191623.
RA FANG G.W., ALLISON R.F., ZAMBOLIM E.M., MAXWELL D.P., GILBERTSON R.L.;
RT "The complete nucleotide sequence and genome organization of bean
RT common mosaic virus (NL3 strain).";
RL VIRUS RES. 39:13-23(1995).
CC -1- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID
CC TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.
CC -1- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
CC MAY BE INVOLVED IN REPLICATION.
CC -1- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.
CC -1- PTM: VPC IS COVALENTLY LINKED TO THE GENOMIC RNA.
CC -1- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE
CC POLYPEPTIDE WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
CC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
CC INDIVIDUAL PROTEINS.
CC -1- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
CC -1- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
CC -1- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U19287; G1373002;
DR PFAM; PF00271; helicase_C; 1.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00767; Poty_coat; 1.
DR PFAM; PF00851; Peptidase_C6; 1.
DR PFAM; PF00863; Peptidase_C4; 1.
KW HYDROLASE; TRANSFERASE; THIOL PROTEASE; RNA-DIRECTED RNA POLYMERASE;
KW COAT PROTEIN; POLYPROTEIN; COVALENT PROTEIN-RNA LINKAGE; HELICASE;
KW ATP-BINDING.
FT CHAIN 1 ? ? N-TERMINAL PROTEIN.
FT CHAIN ? ? ? HELPER COMPONENT PROTEINASE.
FT CHAIN ? ? ? PROTEIN P3.
FT CHAIN ? ? ? 6 KD PROTEIN 1.
FT CHAIN ? ? ? CYTOPLASMIC INCLUSION PROTEIN.
FT CHAIN ? ? ? 6 KD PROTEIN 2.
FT CHAIN ? ? ? GENOME-LINKED PROTEIN.
FT CHAIN ? ? ? NUCLEAR INCLUSION PROTEIN A.
FT CHAIN ? ? ? NUCLEAR INCLUSION PROTEIN B.
FT CHAIN ? ? ? COAT PROTEIN.
FT NP_BIND 1258 1265 ATP (POTENTIAL)
SQ SEQUENCE 3066 AA; 350387 MW; 67B875D6 CRC32;

Query Match 66.2%; Score 45; DB 1; Length 3066;
Best Local Similarity 54.5%; Pred. No. 5.97e+00;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 1194 STKESSTSF 1204
QY 81 STSNDTTSAAF 91
||:||||:|

Search completed: Wed Sep 1 16:31:25 1999
Job time : 11 secs.

```

CC SACCHAROMYCETACEAE; SACCHAROMYCES.

CC [1]

CC SEQUENCE FROM N.A.

CC DIETRICH F.S. / AB972;

CC RA AVILES E., BERNI A., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,

CC RA CHYAN R., DUNCAN M., GUZMAN E., HARTZELL G., HUNICK-SMITH S.,

CC RA HUANG R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,

CC RA MOSEDALE D., NAKAHARA K., NAMATH A., NORGREN R., OEFNER P., OH C.,

CC RA PETEL F.X., ROBERTS D., SEHL P., SCHRAMM S., SHOENEN T., SMITH V.,

CC RA TAYLOR P., WEI Y., YELTON M., BOTSTEIN D., DAVIS R.W.,

CC RA SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.

CC CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC CC

CC EMBL; U18796; G603266; -

CC DR HYPOTHETICAL PROTEIN.

CC KW

CC SEQUENCE 1076 AA; 119349 MW; 9FED783D CRC32;

CC

CC Query Match 67.68; Score 46; DB 1; Length 1076;

CC Best Local Similarity 50.08; Pred. No. 3.44e+00;

CC Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

CC

CC Db 97 TSTNQTSTNSFV 108

CC :::::||||:|

CC Qy 81 TSTNQTSTSAFV 92

CC

CC RESULT 13

CC ID PURA_METJA STANDARD; PRT: 345 AA.

CC AC Q57981;

CC DT 01-NOV-1997 (REL. 35, CREATED)

CC DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

CC DE ADENYLOSUCCINATE SYNTHETASE (EC 6.3.4.4) (IMP--ASPARTATE LIGASE).

CC DE ADENYLOSUCCINATE SYNTHETASE (EC 6.3.4.4) (IMP--ASPARTATE LIGASE).

CC GN PURA OR M10561

CC OS MATHANOCOCUS JANNASCHII.

CC OS ARCHAEA; EURYARCHAEOTA; MATHANOCOCCEAE; MATHANOCOCCEACEAE;

CC OS MATHANOCOCUS.

CC CC

CC [1]

CC SEQUENCE FROM N.A.

CC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;

CC RC MEDLINE; 96337999.

CC RX

CC BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,

CC RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,

CC RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,

CC RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,

CC RA SCOTT J.L., GEORGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.-L., NGUYEN D.,

CC RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOM P.W., HANNA M.C.,

CC RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,

CC RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;

CC RT "Complete genome sequence of the methanogenic archaeon, Methanococcus

CC RT jannaschii."

CC RL SCIENCE 273:1058-1073(1996).

CC CC

CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE DE NOVO PATHWAY OF PURINE

CC NUCLEOTIDE BIOSYNTHESIS.

CC CC

CC -!- CATALYTIC ACTIVITY: GTP + IMP + L-ASPARTATE = GDP +

CC ORTHOPHOSPHATE + ADENYLOSUCCINATE.

CC CC

CC -!- PATHWAY: FIRST COMMITTED STEP IN AMP BIOSYNTHESIS.

CC CC

CC -!- SIMILARITY: WITH OTHER ADENYLOSUCCINATE SYNTHETASES.

CC CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: X81479; G784994; -.
MIN; 600493; -.
DR PROSITE: PS00649; G_PROTEIN_RECP_F2_1; FALSE_NEG.
DR PROSITE: PS00650; G_PROTEIN_RECP_F2_2; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 6.
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS01187; EGF_CA; 5.
DR PFAM: PF00008; EGF; 3.
HSP: P00736; IAP0.
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; RECEPTOR; GLYCOPROTEIN;
KW EGF-LIKE DOMAIN; REPEAT; SIGNAL.
FT SIGNAL 1 17
FT CHAIN 18 886
FT DOMAIN 18 886 CELL SURFACE GLYCOPROTEIN EMRI.
FT TRANSMEM 18 599 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 600 627 POTENTIAL.
FT DOMAIN 628 634 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 635 656 POTENTIAL.
FT DOMAIN 657 666 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 667 690 POTENTIAL.
FT DOMAIN 691 709 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 710 731 POTENTIAL.
FT DOMAIN 732 747 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 748 776 POTENTIAL.
FT DOMAIN 777 794 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 795 814 POTENTIAL.
FT DOMAIN 815 829 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 830 852 POTENTIAL.
FT DOMAIN 853 886 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 80 131 EGF-LIKE 1.
FT DOMAIN 132 171 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 172 213 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 221 267 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 268 316 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 317 599 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT DISULFID 35 47 SER/THR-RICH.
FT DISULFID 41 56 BY SIMILARITY.
FT DISULFID 58 78 BY SIMILARITY.
FT DISULFID 84 97 BY SIMILARITY.
FT DISULFID 91 106 BY SIMILARITY.
FT DISULFID 108 130 BY SIMILARITY.
FT DISULFID 136 148 BY SIMILARITY.
FT DISULFID 142 157 BY SIMILARITY.
FT DISULFID 159 170 BY SIMILARITY.
FT DISULFID 176 188 BY SIMILARITY.
FT DISULFID 182 197 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 225 235 BY SIMILARITY.
FT DISULFID 229 244 BY SIMILARITY.
FT DISULFID 246 266 BY SIMILARITY.
FT DISULFID 272 285 BY SIMILARITY.
FT DISULFID 279 294 BY SIMILARITY.
FT DISULFID 296 315 BY SIMILARITY.
FT CARBOHYD 94 94 POTENTIAL.
FT CARBOHYD 99 99 POTENTIAL.
FT CARBOHYD 127 127 POTENTIAL.
FT CARBOHYD 167 167 POTENTIAL.
FT CARBOHYD 189 189 POTENTIAL.
FT CARBOHYD 194 194 POTENTIAL.
FT CARBOHYD 232 232 POTENTIAL.
FT CARBOHYD 258 258 POTENTIAL.
FT CARBOHYD 312 312 POTENTIAL.
FT CARBOHYD 366 366 POTENTIAL.
FT CARBOHYD 375 375 POTENTIAL.
FT CARBOHYD 448 448 POTENTIAL.
FT CARBOHYD 661 661 POTENTIAL.
886 AA; 97680 MW; 36A07647 CRC32;
SEQUENCE

Query Match 70.6%; Score 48; DB 1; Length 886;
Best Local Similarity 61.5%; Pred. No. 1.11e+00;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Db 470 SESTETTGVAFVS 482
QY 81 STSNDTTSAAFVS 93
RESULT 10
ID NIT2_NEUCR STANDARD; PRT; 1036 AA.
AC P19212;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE NITROGEN CATABOLIC ENZYME REGULATORY PROTEIN.
GN NIT-2.
OS NEUROSPORA CRASSA.
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
OC SORDARIALES; SORDARIACEAE; NEUROSPORA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-74-OR23-1A;
RX MEDLINE; 90158568.
RA FU Y.-H., MARZLUF G.A.;
RT "nit-2, the major nitrogen regulatory gene of Neurospora crassa,"
RT encodes a protein with a putative zinc finger DNA-binding domain.";
RL MOL. CELL. BIOL. 10:1056-1063(1990).
RN [2]
RP MUTAGENESIS.
RX MEDLINE; 91186820.
RA FU Y.-H., MARZLUF G.A.;
RT "Site-directed mutagenesis of the 'zinc finger' DNA-binding domain of
the nitrogen-regulatory protein NIT2 of Neurospora.";
RL MOL. MICROBIOL. 4:1847-1852(1990).
CC -1- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. DURING CONDITIONS
OF NITROGEN LIMITATION IT TURNS ON THE EXPRESSION OF GENES FOR
ENZYMES WHICH ARE REQUIRED FOR THE USE OF A VARIETY OF SECONDARY
NITROGEN SOURCES, INCLUDING NITRATES, PURINES, AMINO ACIDS, AND
PROTEINS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- INDUCTION: BY LACK OF A PRIMARY NITROGEN SOURCE.
CC -1- SIMILARITY: HIGH TO OTHER FUNGAL NITROGEN REGULATORY PROTEINS.
CC -1- SIMILARITY: TO GATA TRANSCRIPTION FACTORS IN THE ZINC-FINGER
REGION.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: M33956; G1107471; -.
DR PIR: A34755; A34755.
DR PROSITE: PS00344; GATA_ZN_FINGER; 1.
DR PFAM: PF00320; GATA; 1.
DR HSP: P17429; 5GAT.
DR TRANSFAC; T00627; -.
DR TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING; ZINC-FINGER;
KW NUCLEAR PROTEIN; NITRATE ASSIMILATION; REPEAT.
FT DOMAIN 49 110 3 X APPROXIMATE REPEATS.
FT REPEAT 49 55 1.
FT REPEAT 87 92 2.
FT REPEAT 105 110 3.
FT ZN_FING 743 767 GATA-TYPE.
FT MUTAGEN 743 746 CTNCG->STNG: ABOLISH DNA-BINDING.
FT MUTAGEN 755 756 RR->GG: ABOLISH DNA-BINDING.
FT MUTAGEN 765 766 NA->DV: ABOLISH DNA-BINDING.
FT MUTAGEN 768 769 GL->DV: ABOLISH DNA-BINDING.
FT MUTAGEN 789 790 KR->NS: ABOLISH DNA-BINDING.


```

RESULT 7
ID CCB2_RABIT STANDARD; PRT; 632 AA.
AC PS4298;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CALCIUM CHANNEL BETA-2 SUBUNIT
DE (CAB2).
GN CACB2 OR CACNLB2.
OS ORYCTOLAGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE; 92192022.
RX HULLIN R., SINGER-LAHAT D., FREICHEL M., BIEL M., DASCAL N.,
RA HOFMANN F., FLOCKERZI V.
RT "Calcium channel beta subunit heterogeneity: functional expression of
RT cloned cDNA from heart, aorta and brain.";
RL EMBO J. 11:885-890(1992)
CC -1- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROPYRIDINE (DHP)
CC SENSITIVE CALCIUM CHANNEL.
CC -1- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:
CC ALPHA-1, ALPHA-2, BETA AND GAMMA.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN HEART, AORTA AND
CC BRAIN.
CC -1- ALTERNATIVE PRODUCTS: THREE FORMS OF THE BETA-2 SUBUNIT ARE
CC PRODUCED BY ALTERNATIVE SPLICING OF THE GENE: CAB2A, CAB2B (SHOWN
CC HERE) AND CAB2C.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64298; G1500; -.
DR EMBL; X64297; G1498; -.
DR EMBL; X64299; G1502; -.
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00774; Ca_channel_B; 1.
KW IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;
KW GLYCOPROTEIN; PHOSPHORYLATION; SH3 DOMAIN; MULTIGENE FAMILY;
KW ALTERNATIVE SPLICING.
FT DOMAIN 86 147 SH3.
FT CARBOHYD 94 94 POTENTIAL.
FT CARBOHYD 175 175 POTENTIAL.
FT VARSPPLIC 1 43 MNOASGLDLKISYKGRKRNFKSGDSTSDTTSNSFV
FT FT 197 234 RQ -> MLDRLAAPHQGLVLE (IN CAB2A).
FT VARSPPLIC 235 240 MISSING (IN CAB2C).
FT FT 235 240 MPFEKK -> KQKQKS (IN CAB2C).
SQ SEQUENCE 632 AA; 70943 MW; DF97DA5D CRC32;

Query Match 79.4%; Score 54; DB 1; Length 632;
Best Local Similarity 75.0%; Pred. No. 3.03e-02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 30 STSDTTSNSFV 41
|:|:|:|:|:|:|
QY 81 STSDNTTSAAFV 92

RESULT 8
ID LPFD_SALTY STANDARD; PRT; 355 AA.
AC P43663;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE LPFD PROTEIN PRECURSOR.
GN LPFD.
OS SALMONELLA TYPHIMURIUM.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 14028;
RX MEDLINE; 95238281.
RA BAUEMLER A.J., HEFFRON F.;
RT "Identification and sequence analysis of lpfABCDE, a putative
RT fimbrial operon of Salmonella typhimurium.";
RL J. BACTERIOL. 177:2087-2097(1995).
CC -1- SIMILARITY: BELONGS TO THE FIMH / LPFD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U18559; G829374; -.
DR STYGENE; SG10560; LPFD.
DR PFAM; PF00419; Fimbrial; 1.
KW FIMBRIA; SIGNAL.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 355 LPFD PROTEIN.
SQ SEQUENCE 355 AA; 37714 MW; 20D529B5 CRC32;

Query Match 70.6%; Score 48; DB 1; Length 355;
Best Local Similarity 63.6%; Pred. No. 1.11e+00;
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 51 TSKNTTTCATFV 61
|:|:|:|:|:|:|
QY 82 TSDNTTSAAFV 92

RESULT 9
ID EMR1_HUMAN STANDARD; PRT; 886 AA.
AC Q14246;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE CELL SURFACE GLYCOPROTEIN EMR1 PRECURSOR (EMR1 HORMONE RECEPTOR).
GN EMR1
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95324926.
RA BAUD V., CHISSOE S.L., VIEGAS-PEQUIGNOT E., DIRIONG S., N'GUYEN V.C.,
RA ROE B.A., LIPINSKI M.;
RT "EMR1, an unusual member in the family of hormone receptors with
RT seven transmembrane segments.";
RL GENOMICS 26:334-344(1995).
CC -1- FUNCTION: PROBABLY INVOLVED IN CELLULAR RESPONSE TO A HORMONE OR
CC AN INTERACTION WITH A PROTEIN LIGAND.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: WIDE EXPRESSION; INCREASED LEVELS IN
CC PERIPHERAL BLOOD MONONUCLEAR CELLS.
CC -1- PTM: N- AND O-GLYCOSYLATED; (POSSIBLE).
CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```


DIHYDROPYRIDINE-SENSITIVE L-TYPE, CHANNEL BETA-1-B2 SUBUNIT
(BETA-1 ISOFORM A).
GN CACNB1 OR CACNLB1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-HIPOCAMPUS;
RX POWERS P.A., LIU S., HOGAN K., GREGG R.G.;
RT "Skeletal muscle and brain isoforms of a beta-subunit of human
voltage-dependent calcium channels are encoded by a single gene.";
RL J. BIOL. CHEM. 267:22967-22972(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-HEART;
RX MEDLINE; 93265672.
RA COLLIN T., WANG J., NARGEOT J., SCHWARTZ A.;
RT "Molecular cloning of three isoforms of the L-type voltage-dependent
calcium channel beta subunit from normal human heart.";
RL CIRC. RES. 72:1337-1344(1993).
CC -!- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROPYRIDINE (DHP)
SENSITIVE CALCIUM CHANNEL.
CC -!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:
ALPHA-1, ALPHA-2, BETA AND GAMMA.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ASSOCIATED WITH
THE CYTOPLASMIC ASPECT OF THE SARCOLEMMMA.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, HEART AND SPLEEN.
CC -!- ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1
SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; M92303; G179806; -
DR EMBL; L06110; G187015; -
DR PIR; C44461; C44461.
DR MIN; I14207; -
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00774; Ca_channel_B; 1.
KW IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;
GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;
MULTIGENE FAMILY.
FT DOMAIN 100 161
FT CARBOHYD 189 189
FT CARBOHYD 425 425
FT PROSITE; PS50002; SH3; 1.
FT TS -> SG (IN REF. 2).
FT G -> R (IN REF. 2).
FT H -> D (IN REF. 2).
FT KL -> TV (IN REF. 2).
FT G -> S (IN REF. 2).
FT T -> S (IN REF. 2).
FT SNTR -> LQHT (IN REF. 2).
FT I -> L (IN REF. 2).
FT M -> I (IN REF. 2).
FT AA -> RR (IN REF. 1).
FT E -> D (IN REF. 2).
FT M -> V (IN REF. 2).
FT S -> N (IN REF. 2).
FT R -> W (IN REF. 2).
FT L -> P (IN REF. 2).
FT L -> P (IN REF. 2).
FT A -> GTP (IN REF. 2).
FT MISSING (IN REF. 2).

FT CONFLICT 557 557 L -> M (IN REF. 2).
FT CONFLICT 571 572 WP -> CA (IN REF. 2).
FT CONFLICT 591 591 R -> Q (IN REF. 2).
SQ SEQUENCE 596 AA; 65638 MW; OCC447D4 CRC32;
Query Match 79.4%; Score 54; DB 1; Length 596;
Best Local Similarity 75.0%; Pred. No. 3.03e-02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 44 STSDTTNSFV 55
Qy 81 STSNDTTSAFV 92
RESULT 6
ID CCBA.RAT STANDARD; PRT; 597 AA.
AC P54283;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CHANNEL BETA-1-B2 SUBUNIT
(BETA-1 ISOFORM A).
GN CACNB1 OR CACNLB1.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 92038046.
RA PRAGNELL M., SAKAMOTO J., JAY S.D., CAMPBELL K.P.;
RT "Cloning and tissue-specific expression of the brain calcium channel
beta-subunit";
RL FEBS LETT. 291:253-258(1991).
CC -!- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROPYRIDINE (DHP)
SENSITIVE CALCIUM CHANNEL.
CC -!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:
ALPHA-1, ALPHA-2, BETA AND GAMMA.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ASSOCIATED WITH
THE CYTOPLASMIC ASPECT OF THE SARCOLEMMMA.
CC -!- ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1
SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE (BY
SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; X61394; G55894; -
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00774; Ca_channel_B; 1.
KW IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;
GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;
MULTIGENE FAMILY.
FT DOMAIN 100 161
FT CARBOHYD 189 189
FT CARBOHYD 425 425
FT PROSITE; PS50002; SH3; 1.
SQ SEQUENCE 597 AA; 65679 MW; 23B08C47 CRC32;
Query Match 79.4%; Score 54; DB 1; Length 597;
Best Local Similarity 75.0%; Pred. No. 3.03e-02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 44 STSDTTNSFV 55
Qy 81 STSNDTTSAFV 92

CC -!- ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1
 CC SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE.
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL: M92302; G179804; -
 CC EMBL: M76560; G179744; -
 CC EMBL: L06111; G187017; -
 CC EMBL: U86960; G215254; -
 CC EMBL: U86952; G215254; JOINED.
 CC EMBL: U86953; G215254; JOINED.
 CC EMBL: U86954; G215254; JOINED.
 CC EMBL: U86955; G215254; JOINED.
 CC EMBL: U86956; G215254; JOINED.
 CC EMBL: U86957; G215254; JOINED.
 CC EMBL: U86958; G215254; JOINED.
 CC EMBL: U86959; G215254; JOINED.
 CC PIR: B44461; B44461.
 CC MIM: 114207; -
 CC PROSITE: PS50002; SH3; 1.
 CC PFAM: PF00018; SH3; 1.
 CC PFAM: PF00774; Ca_channel_B; 1.
 CC IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;
 CC GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;
 CC MULTIGENE FAMILY.
 CC DOMAIN 100 161 SH3
 CC FT CARBOHYD 189 189 POTENTIAL.
 CC FT CARBOHYD 425 425 POTENTIAL.
 CC FT CONFLICT 21 21 E -> G (IN REF. 2).
 CC FT CONFLICT 28 29 QG -> R (IN REF. 3).
 CC FT CONFLICT 434 435 AA -> RR (IN REF. 1 AND 4).
 CC SQ SEQUENCE 478 AA; 53006 MW; 6011892D CRC32;
 CC
 CC Query Match 79.4%; Score 54; DB 1; Length 478;
 CC Best Local Similarity 75.0%; Pred. No. 3.03e-02;
 CC Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC Db 44 STSSDTSNSPV 55
 CC QY 81 STSNDTTSAPV 92
 CC
 CC
 CC RESULT 2
 CC ID CCBC_HUMAN STANDARD; PRT; 523 AA.
 CC AC Q02639;
 CC DT 01-JUN-1994 (REL. 29, CREATED)
 CC DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 CC DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 CC DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CALCIUM CHANNEL BETA-1M SUBUNIT
 CC DE (BETA-1 ISOFORM C) (BETA-1A).
 CC GN CACNB1 OR CACNLB1.
 CC OS HOMO SAPIENS (HUMAN).
 CC OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 CC OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 CC RN [1]
 CC RN SEQUENCE FROM N.A.
 CC RN TISSUE-SKELETAL MUSCLE;
 CC RX MEDLINE: 93054616
 CC RA POWERS P.A., LIU S., HOGAN K., GREGG R.G.;
 CC RT "Skeletal muscle and brain isoforms of a beta-subunit of human
 CC RT voltage-dependent calcium channels are encoded by a single gene."
 CC RL J. BIOL. CHEM. 267:22967-22972(1992).
 CC RN [2]
 CC RN SEQUENCE FROM N.A.
 CC RP SEQUENCE HEART;
 CC RC

RX MEDLINE: 93265672.
 RA COLLIN T., WANG J., NARGEOT J., SCHWARTZ A.;
 RT "Molecular cloning of three isoforms of the L-type voltage-dependent
 RT calcium channel beta subunit from normal human heart."
 RL CIRC. RES. 72:1337-1344(1993).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP POWERS P.A., GREGG R.G., HOGAN K.;
 RA SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [4]
 RN SEQUENCE OF 146-209 FROM N.A.
 RX MEDLINE: 93372845.
 RA ILES D.E., SEGERS B., SENGERS R.C.A., MONSIEURS K., HEYTENS L.,
 RA HALSALL P.J., HOPKINS P.M., ELLIS F.R., HALL-CURRAN J.L.,
 RA STEWART A.D., WIERINGA B.;
 RT "Genetic mapping of the beta 1- and gamma-subunits of the human
 RT skeletal muscle L-type voltage-dependent calcium channel on
 RT chromosome 17q and exclusion as candidate genes for malignant
 RT hyperthermia susceptibility."
 RL HUM. MOL. GENET. 2:863-868(1993).
 CC -!- FUNCTION: THIS PROTEIN IS THE BETA SUBUNIT OF THE DIHYDROPYRIDINE
 CC (DHP) SENSITIVE CALCIUM CHANNEL. PLAYS A ROLE IN EXCITATION-
 CC CONTRACTION COUPLING. THE SKELETAL MUSCLE DHP-SENSITIVE CA(2+)
 CC CHANNEL MAY FUNCTION ONLY AS A MULTIPLE SUBUNIT COMPLEX.
 CC -!- SUBUNIT: THE SKELETAL MUSCLE L-TYPE CALCIUM CHANNEL IS COMPOSED OF
 CC FIVE SUBUNITS: ALPHA-1, ALPHA-2, BETA, GAMMA AND DELTA.
 CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ASSOCIATED WITH
 CC THE CYTOPLASMIC ASPECT OF THE SARCOLEMA.
 CC -!- TISSUE SPECIFICITY: SKELETAL MUSCLE.
 CC -!- ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1
 CC SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE.
 CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL: M92301; G179802; -
 CC EMBL: L06112; G187019; -
 CC EMBL: U86960; G215255; -
 CC EMBL: U86952; G215255; JOINED.
 CC EMBL: U86953; G215255; JOINED.
 CC EMBL: U86954; G215255; JOINED.
 CC EMBL: U86955; G215255; JOINED.
 CC EMBL: U86956; G215255; JOINED.
 CC EMBL: U86957; G215255; JOINED.
 CC EMBL: U86958; G215255; JOINED.
 CC EMBL: U86959; G215255; JOINED.
 CC EMBL: Z21725; G38563; -
 CC EMBL: Z21726; G38565; -
 CC MIM: 114207; -
 CC PROSITE: PS50002; SH3; 1.
 CC PFAM: PF00018; SH3; 1.
 CC PFAM: PF00774; Ca_channel_B; 1.
 CC IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;
 CC GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;
 CC MULTIGENE FAMILY.
 CC DOMAIN 100 161 SH3
 CC FT CARBOHYD 189 189 POTENTIAL.
 CC FT CARBOHYD 470 470 POTENTIAL.
 CC FT CONFLICT 28 29 QG -> R (IN REF. 2).
 CC FT CONFLICT 183 183 MISSING (IN REF. 4).
 CC FT CONFLICT 479 480 AA -> RR (IN REF. 1 AND 3).
 CC SQ SEQUENCE 523 AA; 57863 MW; 32CBE9D CRC32;
 CC
 CC Query Match 79.4%; Score 54; DB 1; Length 523;
 CC Best Local Similarity 75.0%; Pred. No. 3.03e-02;
 CC Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

(TM)

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	54	79.4	478	1	CCBB_HUMAN	DIHYDROPYRIDINE-SENSIT	3.03e-02
2	54	79.4	523	1	CCBC_HUMAN	DIHYDROPYRIDINE-SENSIT	3.03e-02
3	54	79.4	524	1	CCBC_RABIT	DIHYDROPYRIDINE-SENSIT	3.03e-02
4	54	79.4	567	1	CCB2_HUMAN	DIHYDROPYRIDINE-SENSIT	3.03e-02
5	54	79.4	596	1	CCBA_HUMAN	DIHYDROPYRIDINE-SENSIT	3.03e-02
6	54	79.4	597	1	CCBA_RAT	DIHYDROPYRIDINE-SENSIT	3.03e-02
7	54	79.4	632	1	CCB2_RABIT	DIHYDROPYRIDINE-SENSIT	3.03e-02
8	48	70.6	355	1	LPFD_SALTY	LPFD PROTEIN PRECURSOR	3.03e-02
9	48	70.6	886	1	EMT1_HUMAN	CELL SURFACE GLYCOPROT	1.11e+00
10	48	70.6	1036	1	NR12_NEUCR	NETROGEN CATABOLIC ENZ	1.11e+00
11	46	67.6	532	1	CPPL1_DROAC	CYTCHROME P450 12B1 P	3.44e+00
12	46	67.6	1076	1	YEM3_YEAST	HYPOTHETICAL 119.3 KD	1.11e+00
13	45	66.2	345	1	PUR4_METJA	ADENYLOSUCINATE SYNTH	3.44e+00
14	45	66.2	416	1	IR12_HCMVA	HYPOTHETICAL PROTEIN I	5.97e+00
15	45	66.2	3066	1	POLG_BCMVN	GENOME POLYPROTEIN [CO	5.97e+00
16	44	64.7	490	1	C1KL_DROME	VOLTAGE-GATED POTASSIU	1.03e+01
17	44	64.7	588	1	PR28_YEAST	PRE-MRNA SPLICING FACT	1.03e+01
18	44	64.7	607	1	HRAL_XANCV	HYPERSENSITIVITY RESPO	1.03e+01
19	44	64.7	3106	1	LMA2_MOUSE	LAMININ ALPHA-2 CHAIN	1.03e+01
20	44	64.7	3110	1	LMA2_HUMAN	LAMININ ALPHA-2 CHAIN	1.03e+01
21	43	63.2	142	1	RPC9_YEAST	DNA-DIRECTED RNA POLYM	1.75e+01
22	43	63.2	269	1	FOLD_MYCPN	METHYLENETETRAHYDROFOL	1.75e+01
23	43	63.2	337	1	YEB7_YEAST	HYPOTHETICAL 38.2 KD P	1.75e+01

```
#submission submitted to the EMBL Data Library, December 1994
#description The sequence of S. cerevisiae cosmids 9379, 9581, and lambda
clone 4678.
#accession S50536
##molecule_type DNA
##residues 1-1076 #label D1E
##cross-references EMBL:U18796; NID:9603265; PID:9603266; MIPS:YER033c
GENETICS
#map_position 5R
SUMMARY #length 1076 #molecular-weight 119349 #checksum 9442
Query Match 67.6%; Score 46; DB 2; Length 1076;
Best Local Similarity 50.0%; Pred. No. 9.97e+00;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 97 TSTNQITTSNSFV 108
QY 81 STSNDTTSAAFV 92
:::||||:|
:::||||:|

Search completed: Wed Sep 1 16:32:49 1999
Job time : 14 secs.
```

```

|||||:|
Qy 81 STSNDTSAAFV 92

RESULT 11
ENTRY S21048 #type complete
TITLE calcium channel protein beta chain CaB2b - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
DATE 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change
10-Sep-1997

ACCESSIONS S21048
REFERENCE S21046
#authors Hulin, R.; Singer-Lahat, D.; Freichel, M.; Biel, M.; Dascal,
N.; Hofmann, F.; Flockerzi, V.
#journal EMBO J. (1992) 11:885-890
#title Calcium channel beta subunit heterogeneity: functional
expression of cloned cDNA from heart, aorta and brain.
#cross-references M01b:92192022
#accession S21048
#molecule_type mRNA
#residues 1-632 #label HUL
#cross-references EMBL:X64298; NID:g1499; PID:g1500
SUMMARY #length 632 #molecular-weight 70943 #checksum 1583

Query Match 79.4%; Score 54; DB 2; Length 632;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 30 STSDDTNSFV 41
|||||:|
Qy 81 STSNDTSAAFV 92

RESULT 12
ENTRY D56271 #type complete
TITLE long polar fimbrial operon protein lpfd - Salmonella
typhimurium
ORGANISM #formal_name Salmonella typhimurium
DATE 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change
09-Sep-1997

ACCESSIONS D56271
REFERENCE A56271
#authors Baumber, A.J.; Heffron, F.
#journal J. Bacteriol. (1995) 177:2087-2097
#title Identification and sequence analysis of lpfABCD, a putative
fimbrial operon of Salmonella typhimurium.
#accession D56271
#status Preliminary
#molecule_type DNA
#residues 1-355 #label BAE
#cross-references GB:U18559; NID:g829370; PID:g829374
GENETICS lpfd
#gene
SUMMARY #length 355 #molecular-weight 37714 #checksum 5448

Query Match 70.6%; Score 48; DB 2; Length 355;
Best Local Similarity 63.6%; Pred. No. 3.61e+00;
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 51 TSKNTTGATFV 61
|||||:|
Qy 82 TSNDTSAAFV 92

RESULT 13
ENTRY A57172 #type complete
TITLE probable hormone receptor EMR1 precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change
24-Sep-1998
ACCESSIONS A57172
REFERENCE A57172

```

```

#authors Baud, V.; Chissoe, S.L.; Viegas-Pequignot, E.; Diriong, S.;
N'Guyen, V.C.; Roe, B.A.; Lipinski, M.
#journal Genomics (1995) 26:334-344
#title EMR1, an unusual member in the family of hormone receptors
with seven transmembrane segments.
#accession A57172
#status preliminary
#molecule_type mRNA
#residues 1-886 #label BAU
#cross-references GB:X81479; NID:g784993; PID:g784994
GENETICS
#gene GDB:EMR1
#cross-references GDB:378349; OMIM:600493
#map_position 19p13.3-19p13.3
KEYWORDS transmembrane protein
SUMMARY #length 886 #molecular-weight 97679 #checksum 2055

Query Match 70.6%; Score 48; DB 2; Length 886;
Best Local Similarity 61.5%; Pred. No. 3.61e+00;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 470 SESTETGAFVS 482
|||||:|
Qy 81 STSNDTSAAFV 93

RESULT 14
ENTRY A34755 #type complete
TITLE nitrogen regulatory protein nit-2 - Neurospora crassa
ORGANISM #formal_name Neurospora crassa
DATE 13-Jul-1990 #sequence_revision 26-Jul-1996 #text_change
16-Feb-1997

ACCESSIONS A34755
REFERENCE A34755
#authors Fu, Y.H.; Marzluf, G.A.
#journal Mol. Cell. Biol. (1990) 10:1056-1065
#title nit-2, the major nitrogen regulatory gene of Neurospora
crassa, encodes a protein with a putative zinc finger
DNA-binding domain.
#cross-references M01d:90158568
#accession A34755
#molecule_type DNA; mRNA
#residues 1-1036 #label FUY
#cross-references GB:M33956
GENETICS
#introns 209/2; 335/3
CLASSIFICATION #superfamily nitrogen regulatory protein nit-2; GATA-type
zinc finger homology
KEYWORDS DNA binding; transcription regulation; zinc finger
FEATURE 740-793 #domain GATA-type zinc finger homology #label GZF\
743-767 #region zinc finger GATA motif
SUMMARY #length 1036 #molecular-weight 109295 #checksum 235

Query Match 70.6%; Score 48; DB 1; Length 1036;
Best Local Similarity 53.8%; Pred. No. 3.61e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 87 SSSNNNTSNGFVA 99
|||||:|
Qy 81 STSNDTSAAFV 93

RESULT 15
ENTRY S50536 #type complete
TITLE hypothetical protein YER033c - yeast (Saccharomyces
cerevisiae)
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 28-May-1993 #sequence_revision 24-Feb-1995 #text_change
21-Nov-1997
ACCESSIONS S50536
REFERENCE S50428
#authors Dietrich, F.S.

```

```
REFERENCE      A41347
#authors      Ruth, P.; Rohrkasten, A.; Biel, M.; Bosse, E.; Regulla, S.;
#journal      Meyer, H.E.; Flockerzi, V.; Hofmann, F.
#title        Science (1989) 245:1115-1118
#keywords      Primary structure of the beta subunit of the DHP-sensitive
               calcium channel from skeletal muscle.
#cross-references MUID:89368946
#accession     A41347
#status        preliminary
#molecule_type mRNA
#residues      1-524 #label RUT
#cross-references GB:M25817
#keywords      skeletal muscle
#summary       #length 524 #molecular-weight 57868 #checksum 8438

Query Match      79.4%; Score 54; DB 2; Length 524;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches          9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTSNSFV 55
QY 81 STSNDTTSAAVF 92
    |||:||||:|

RESULT 7
ENTRY  A48895 #type complete
TITLE  myasthenic syndrome antigen B - human
ALTERNATE_NAMES  MYSB
ORGANISM #formal_name Homo sapiens #common_name man
DATE      07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change
          20-Mar-1998
ACCESSIONS  A48895
REFERENCE    Rosenfeld, M.R.; Wong, E.; Dalmau, J.; Manley, G.; Posner,
#authors     J.B.; Sher, E.; Furneaux, H.M.
#journal     Ann. Neurol. (1993) 33:113-120.
#title       Cloning and characterization of a Lambert-Eaton myasthenic
               syndrome antigen.
#cross-references MUID:93263585
#accession   A48895
#status      preliminary
#molecule_type mRNA
#residues    1-567 #label ROS
#cross-references GB:S60415; NID:g300416; PID:g300417
#experimental_source fetal brain
#note        sequence extracted from NCBI backbone (NCBIN:132135,
               NCBIP:132136)
KEYWORDS      alternative splicing
#summary      #length 567 #molecular-weight 62087 #checksum 9649

Query Match      79.4%; Score 54; DB 2; Length 567;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches          9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 58 STSDTTSNSFV 69
QY 81 STSNDTTSAAVF 92
    |||:||||:|

RESULT 8
ENTRY  C44461 #type complete
TITLE  voltage-dependent calcium channel beta subunit beta1B2
ORGANISM isoform - human
DATE      10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change
          20-Mar-1998
ACCESSIONS  C44461
REFERENCE    Powers, P.A.; Liu, S.; Hogan, K.; Gregg, R.G.
#authors     J. Biol. Chem. (1992) 267:22967-22972
#journal     Skeletal muscle and brain isoforms of a beta-subunit of human
#title       voltage-dependent calcium channels are encoded by a single
               gene.
```

```
#cross-references MUID:93054616
#accession     C44461
#status        preliminary; not compared with conceptual translation
#molecule_type nucleic acid
#residues      1-596 #label POW
#cross-references GB:M92303; NID:g291880; PID:g179806
#experimental_source hippocampus
#note          sequence extracted from NCBI backbone (NCBIP:118133)
#summary       #length 596 #molecular-weight 65808 #checksum 4540

Query Match      79.4%; Score 54; DB 2; Length 596;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches          9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTSNSFV 55
QY 81 STSNDTTSAAVF 92
    |||:||||:|

RESULT 9
ENTRY  S18304 #type complete
TITLE  calcium channel protein beta chain - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE      22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
          20-Mar-1998
ACCESSIONS  S18304
REFERENCE    Pragnell, M.; Sakamoto, J.; Jay, S.D.; Campbell, K.P.
#authors     FEBS Lett. (1991) 291:253-258
#journal     Cloning and tissue-specific expression of the brain calcium
#title       channel beta-subunit.
#cross-references MUID:92038046
#accession   S18304
#molecule_type mRNA
#residues    1-597 #label PRA
#cross-references GB:X61394; NID:g55893; PID:g55894
#summary      #length 597 #molecular-weight 65679 #checksum 8572

Query Match      79.4%; Score 54; DB 2; Length 597;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches          9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTSNSFV 55
QY 81 STSNDTTSAAVF 92
    |||:||||:|

RESULT 10
ENTRY  I52859 #type complete
TITLE  L-type voltage-gated calcium channel B subunit - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE      02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
          02-Jul-1996
ACCESSIONS  I52859
REFERENCE    Collin, T.; Wang, J.; Margeot, J.; Schwartz, A.
#authors     Circ. Res. (1993) 72:1337-1344
#journal     Molecular cloning of three isoforms of the L-type
#title       voltage-dependent calcium channel B subunit from normal
               human heart.
#cross-references MUID:93265672
#accession   I52859
#status      preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues    1-597 #label RES
#cross-references GB:L06110; NID:g187014; PID:g187015
#summary      #length 597 #molecular-weight 65578 #checksum 5631

Query Match      79.4%; Score 54; DB 2; Length 597;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches          9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTSNSFV 55
```

```
RESULT 2
ENTRY 165766 #type complete
TITLE L-type voltage-gated calcium channel B subunit - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
12-Aug-1996
ACCESSIONS 165766
REFERENCE Collin, T.; Wang, J.; Nargeot, J.; Schwartz, A.
#authors Circ. Res. (1993) 72:1337-1344
#journal Molecular cloning of three isoforms of the L-type
#title voltage-dependent calcium channel B subunit from normal
human heart.
#cross-references MUID:93265672
#accession 165766
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-477 #label RES
#cross-references GB:L06111; NID:gl87016; PID:gl87017
SUMMARY #length 477 #molecular-weight 52977 #checksum 4957
Query Match 79.4%; Score 54; DB 2; Length 477;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 43 STSDTTNSFV 54
||||| :||
QY 81 STSNDTTSAAFV 92

RESULT 3
ENTRY JH0566 #type complete
TITLE calcium channel beta-2 chain - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
10-Sep-1997
ACCESSIONS JH0566
REFERENCE Williams, M.E.; Feldman, D.H.; McCue, A.F.; Brenner, R.;
#authors Vellicelebi, G.; Ellis, S.B.; Harpold, M.M.
#journal Neuron (1992) 8:71-84
#title Structure and functional expression of alpha1, alpha2, and
beta subunits of a novel human neuronal calcium channel
subtype.
#cross-references MUID:92110010
#accession JH0566
#molecule_type mRNA
#residues 1-478 #label WIL
#cross-references GB:M7560; NID:gl79743; PID:gl79744
#experimental_source hippocampus
COMMENT This protein is a subunit of the voltage dependent calcium channel.
KEYWORDS glycoprotein; phosphoprotein
FEATURE
32,167,209,348,374, #binding_site phosphate (Ser) (covalent) (by protein
450,464 kinase C) #status predicted\
64,201 #binding_site phosphate (Thr) (covalent) (by protein
kinase C) #status predicted\
189,425 #binding_site carbohydrate (Asn) (covalent) #status
predicted\
205 #binding_site phosphate (Thr) (covalent) (by
cAMP-dependent kinase) #status predicted\
SUMMARY #length 478 #molecular-weight 52934 #checksum 6465
Query Match 79.4%; Score 54; DB 2; Length 478;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 44 STSDTTNSFV 55
||||| :||
QY 81 STSNDTTSAAFV 92

RESULT 4
ENTRY B44461 #type complete
TITLE voltage-dependent calcium channel beta subunit beta1B1
ORGANISM #formal_name Homo sapiens #common_name man
DATE 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change
20-Mar-1998
ACCESSIONS B44461
REFERENCE Powers, P.A.; Liu, S.; Hogan, K.; Gregg, R.G.
#authors J. Biol. Chem. (1992) 267:22967-22972
#journal Skeletal muscle and brain isoforms of a beta-subunit of human
#title voltage-dependent calcium channels are encoded by a single
gene.
#cross-references MUID:93054616
#accession B44461
#status preliminary
#molecule_type nucleic acid
#residues 1-478 #label POW
#cross-references GB:M92302; NID:gl79803; PID:gl79804
#experimental_source hippocampus
#note sequence inconsistent with the nucleotide translation
#note sequence extracted from NCBI backbone (NCBIP:118131)
SUMMARY #length 478 #molecular-weight 53176 #checksum 7630
Query Match 79.4%; Score 54; DB 2; Length 478;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 44 STSDTTNSFV 55
||||| :||
QY 81 STSNDTTSAAFV 92

RESULT 5
ENTRY I65767 #type complete
TITLE L-type voltage-gated calcium channel B subunit - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
12-Aug-1996
ACCESSIONS I65767
REFERENCE Collin, T.; Wang, J.; Nargeot, J.; Schwartz, A.
#authors Circ. Res. (1993) 72:1337-1344
#journal Molecular cloning of three isoforms of the L-type
#title voltage-dependent calcium channel B subunit from normal
human heart.
#cross-references MUID:93265672
#accession I65767
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-522 #label RES
#cross-references GB:L06112; NID:gl87018; PID:gl87019
SUMMARY #length 522 #molecular-weight 57834 #checksum 7462
Query Match 79.4%; Score 54; DB 2; Length 522;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 43 STSDTTNSFV 54
||||| :||
QY 81 STSNDTTSAAFV 92

RESULT 6
ENTRY A41347 #type complete
TITLE calcium channel protein beta chain,
ORGANISM dihydropyridine-sensitive, skeletal muscle - rabbit
#formal_name Oryctolagus cuniculus #common_name domestic
#title rabbit
DATE 03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change
07-Feb-1997
ACCESSIONS A41347
```


Result No.	Query		Length	DB	ID	Description	Pred. No.
	Score	Match					
1	54	79.4	475	2	A44461	voltage-dependent cal	1.44e-01
2	54	79.4	477	2	I55766	L-type voltage-gated	1.44e-01
3	54	79.4	478	2	JH0566	calcium channel beta-	1.44e-01
4	54	79.4	472	2	B44461	voltage-dependent cal	1.44e-01
5	54	79.4	522	2	I65767	L-type voltage-gated	1.44e-01
6	54	79.4	524	2	A41347	calcium channel prote	1.44e-01
7	54	79.4	567	2	A48895	myasthenic syndrome a	1.44e-01
8	54	79.4	596	2	C44461	voltage-dependent cal	1.44e-01
9	54	79.4	597	2	S18304	calcium channel prote	1.44e-01
10	54	79.4	597	2	I52859	L-type voltage-gated	1.44e-01
11	54	79.4	632	2	S21048	calcium channel prote	1.44e-01
12	48	70.6	355	2	D56271	long polar flmbrial o	3.61e+00
13	48	70.6	886	2	A57172	probable hormone rece	3.61e+00
14	48	70.6	1036	1	A34755	nitrogen regulatory p	3.61e+00
15	46	67.6	1076	2	S50536	hypothetical protein	9.97e+00
16	45	66.2	345	2	A64370	adenylosuccinate synt	1.64e+01
17	45	66.2	416	2	S09761	hypothetical protein	1.64e+01
18	45	66.2	583	2	S50959	probable membrane pro	1.64e+01
19	44	64.7	37	2	A21112	variant surface glyco	2.67e+01
20	44	64.7	336	2	I57681	potassium channel pro	2.67e+01
21	44	64.7	490	2	A35312	potassium channel pro	2.67e+01
22	44	64.7	588	2	A39624	probable helicase (PC	2.67e+01
23	44	64.7	1751	1	MNHUWH	laminin alpha-2 chain	2.67e+01

```

OS Homo sapiens.
PN US5792846-A.
PD 11-AUG-1998.
PF 31-MAY-1995. 455543.
PR 04-APR-1994; US-223305.
PR 04-APR-1988; US-176899.
PR 04-APR-1989; US-603751.
PR 04-APR-1989; WO-001408.
PR 20-FEB-1990; US-482384.
PR 30-NOV-1990; US-620250.
PR 15-AUG-1991; US-745206.
PR 31-MAY-1995; US-455543.
PA (SIBI-) SIBIA NEUROSCIENCES INC.
PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
PI Williams ME;
DR WPI: 98-456192/39.
DR N-PSDB: V42688.
PT DNA encoding human calcium channel alpha 1B subunit protein -
PT useful for recombinant production of the channel for screening of
PT its modulators, and diagnosis of Lambert Eaton Syndrome
PS Claim 3; Columns 219-224; 166pp; English.
CC The present sequence represents a splice variant of the beta subunit of
CC a human calcium channel. Calcium channels are membrane-spanning,
CC multi-subunit proteins that allow controlled entry of calcium ions into
CC cells. This leads to depolarisation events required for muscle
CC contraction. The recombinant subunit, when expressed with nucleic acids
CC encoding the complete calcium channel, can be used in assays for the
CC detection and characterisation of compounds that modulate the channel.
CC The DNA encoding the subunits can be alternatively spliced when
CC transcribed, giving more than one form of the protein from the same
CC transcript, each having slightly different properties. In addition, the
CC reactivity of the alpha 1 subunit with IgG molecules from the serum of
CC an individual with Lambert Eaton Syndrome (LES) can be used as a
CC diagnostic for the disease.
SQ Sequence 598 AA;

Query Match 79.4%; Score 54; DB 33; Length 598;
Best Local Similarity 75.0%; Pred. No. 1.78e+01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnfsv 55
Qy 81 STSNDTTSAAVF 92
|||||:||

RESULT 15
ID R72612 standard; Protein; 598 AA.
AC R72612;
DT 01-DEC-1995 (first entry)
DE Human calcium channel subunit beta 1-3.
KW Calcium channel subunit; antagonist; agonist; diagnosis;
KW Lambert Eaton Syndrome.
OS Homo sapiens.
PN WO9504822-A.
PD 16-FEB-1995.
PF 11-AUG-1994; U09230.
PR 11-AUG-1993; US-105536.
PR 05-NOV-1993; US-149097.
PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.
PI Ellis SB, Gillespie A, Harpold MM, McCue AF, Williams ME;
DR WPI: 95-090900/12.
DR N-PSDB: Q87839.
PT DNA encoding human calcium channel sub-unit(s) - used for
PT developing prods. for studying calcium channels, e.g. for
PT obtaining agonists and antagonists
PS Disclosure; Page 163-165; 285pp; English.
CC To isolate DNA encoding the beta 1 subunit, a human hippocampus
CC cDNA library was screened by hybridisation to a DNA fragment
CC encoding a rabbit skeletal muscle calcium channel beta subunit.
CC A hybridising clone was selected and was in turn used to isolate
CC overlapping clones until the overlapping clones encompassing DNA
CC encoding the entire human calcium channel beta 2 subunit were
CC isolated and sequenced. Five alternatively spliced forms of the

```

```

CC beta 1 subunit have been identified. These forms are designated
CC beta 1-1, expressed in skeletal muscle, beta 1-2, expressed in the
CC CNS, beta 1-3, also expressed in the CNS, beta 1-4, expressed in
CC aorta tissue and HEK 293 cells, and beta 1-5, expressed in HEK 293
CC cells. Full-length DNA clones encoding the beta 1-2 and beta 1-3
CC subunits have been constructed. The subunits beta 1-1, beta 1-2,
CC beta 1-4 and beta 1-5 have been identified by nucleic acid
CC amplification, analysis as alternatively spliced forms of the
CC beta subunit. Sequences of the beta 1 splice variants are set
CC forth in Q87838/R72611, Q87839/R72612 and Q87831-Q87833 and
CC R72604-R72706.
SQ Sequence 598 AA;

Query Match 79.4%; Score 54; DB 14; Length 598;
Best Local Similarity 75.0%; Pred. No. 1.78e+01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnfsv 55
Qy 81 STSNDTTSAAVF 92
|||||:||

Search completed: Wed Sep 1 16:33:43 1999
Job time : 36 secs.

```

KW Beta-1 subunit; human; calcium channel; assay; detection;
 KW characterisation; Lambert Eaton Syndrome; LES; diagnosis.
 OS Homo sapiens.
 PN US5792846-A.
 PD 11-AUG-1998.
 PF 31-MAY-1995; 455543.

PR 04-APR-1994; US-223305.
 PR 04-APR-1988; US-176899.
 PR 04-APR-1989; US-603751.
 PR 04-APR-1989; WO-U01408.
 PR 20-FEB-1990; US-482384.
 PR 30-NOV-1990; US-620250.
 PR 15-AUG-1991; US-745206.
 PR 31-MAY-1995; US-455543.

(SIBI-) SIBIA NEUROSCIENCES INC.

PA Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,

PI Williams ME;

DR WPI; 98-456192/39.

DR N-PSDB; V42705.

PT DNA encoding human calcium channel alpha 1B sub:unit protein -
 PT useful for recombinant production of the channel for screening of
 PT its modulators, and diagnosis of Lambert Eaton Syndrome

PS Claim 3; Columns 223-226; 166pp; English.

CC The present sequence represents the beta-1 subunit of a human calcium
 CC channel. Calcium channels are membrane-spanning, multi-subunit proteins
 CC that allow controlled entry of calcium ions into cells. This leads
 CC to depolarisation events required for muscle contraction. The recombinant
 CC subunit, when expressed with nucleic acids encoding the complete calcium
 CC channel, can be used in assays for the detection and characterisation of
 CC compounds that modulate the channel. The DNA encoding the subunits can
 CC be alternatively spliced when transcribed, giving more than one form of
 CC the protein from the same transcript, each having slightly different
 CC properties. In addition, the reactivity of the alpha 1 subunit with IgG
 CC molecules from the serum of an individual with Lambert Eaton Syndrome
 CC (LES) can be used as a diagnostic for the disease.
 SQ Sequence 523 AA;

Query Match 79.4%; Score 54; DB 33; Length 523;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdttsnsfv 55

QY 81 STSNDTTSAAVF 92

RESULT '12

ID W63147 standard; Protein; 530 AA.

AC W63147;

DT 12-OCT-1998 (first entry)

DE Human calcium channel beta subunit.

KW Beta subunit; human; calcium channel; assay; detection;

KW characterisation; Lambert Eaton Syndrome; LES; diagnosis.

OS Homo sapiens.

PN US5792846-A.

PD 11-AUG-1998.

PF 31-MAY-1995; 455543.

PR 04-APR-1994; US-223305.

PR 04-APR-1988; US-176899.

PR 04-APR-1989; US-603751.

PR 04-APR-1989; WO-U01408.

PR 20-FEB-1990; US-482384.

PR 30-NOV-1990; US-620250.

PR 15-AUG-1991; US-745206.

PR 31-MAY-1995; US-455543.

(SIBI-) SIBIA NEUROSCIENCES INC.

PA Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,

PI Williams ME;

DR WPI; 98-456192/39.

DR N-PSDB; V42693.

PT DNA encoding human calcium channel alpha 1B sub:unit protein -

PT useful for recombinant production of the channel for screening of

PT its modulators, and diagnosis of Lambert Eaton Syndrome

PS Disclosure; Columns 125-130; 166pp; English.

CC The present sequence represents the beta subunit of a human calcium
 CC channel. Calcium channels are membrane-spanning, multi-subunit
 CC proteins that allow controlled entry of calcium ions into cells.

CC This leads to depolarisation events required for muscle contraction.

CC The recombinant subunit, when expressed with nucleic acids encoding

CC the complete calcium channel, can be used in assays for the detection

CC and characterisation of compounds that modulate the channel. The

CC DNA encoding the subunits can be alternatively spliced when

CC transcribed, giving more than one form of the protein from the same

CC transcript, each having slightly different properties. In addition, the

CC reactivity of the alpha 1 subunit with IgG molecules from the serum of

CC an individual with Lambert Eaton Syndrome (LES) can be used as a

CC diagnostic for the disease.

SQ Sequence 530 AA;

Query Match 79.4%; Score 54; DB 33; Length 530;

Best Local Similarity 75.0%; Pred. No. 1.78e+01;

Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdttsnsfv 55

QY 81 STSNDTTSAAVF 92

RESULT 13

ID R39565 standard; Protein; 571 AA.

AC R39565;

DT 09-FEB-1994 (first entry)

DE Human neuronal VDCC beta-subunit encoded by clone HBB3.

KW Voltage-dependent calcium channel; VDCC; beta-subunit;

KW calcium flux; ss.

OS Homo sapiens.

PN DE4222156-A.

PD 19-AUG-1993.

PF 06-JUL-1992; 222126.

PR 17-FEB-1992; DE-204716.

PR 06-JUL-1992; DE-222126.

PA (FARB) BAYER AG

PI Spreyer P, Unterbeck A;

DR WPI; 93-265734/34.

DR N-PSDB; Q46078.

PT Human neuronal beta-unit cDNA of voltage dependent calcium

PT channels - useful in calcium-flux studies and screening systems

PT for agonists and antagonists of calcium channels

PS Claim 2; Page 11-13; 13pp; German.

CC A first oligonucleotide probe (Q46075) complementary to nucleotides

CC 361-400 of the VDCC beta-subunit from rabbit skeletal muscle was

CC used to screen a human hippocampus cDNA library. A 1.9kb cDNA

CC fragment was isolated for further screening to isolate human VDCC

CC beta-subunit coding sequences. Clone HBB3 was sequenced and

CC nucleotides 1-1288 were found to have 92% homology to the rabbit

CC beta-subunit; from position 1289 there is no detectable homology.

CC The amino acid sequence R39565 was deduced from the open reading

CC frame. See also Q46076 and Q46077.

SQ Sequence 571 AA;

Query Match 79.4%; Score 54; DB 8; Length 571;

Best Local Similarity 75.0%; Pred. No. 1.78e+01;

Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdttsnsfv 55

QY 81 STSNDTTSAAVF 92

RESULT 14

ID W63144 standard; Protein; 598 AA.

AC W63144;

DT 12-OCT-1998 (first entry)

DE Human calcium channel beta subunit splice variant beta3.

KW Beta subunit; human; calcium channel; assay; detection;

KW characterisation; Lambert Eaton Syndrome; LES; diagnosis.

RESULT 8
 ID R72611 standard; Protein; 478 AA.
 AC R72611.
 DT 01-DEC-1995 (first entry)
 DE Human calcium channel subunit beta 1.
 KW Calcium channel subunit; antagonist; agonist; diagnosis;
 KW Lambert Eaton Syndrome.
 OS Homo sapiens.
 PN W09504822-A.
 PD 16-FEB-1995.
 PF 11-AUG-1994; U09230.
 PR 11-AUG-1993; US-105536.
 PR 05-NOV-1993; US-149097.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Ellis SB, Gillespie A, Harpold MM, Mccue AF, Williams ME;
 DR WPI: 95-090900/12.
 DR N-PSDB: Q87838.
 PT DNA encoding human calcium channel sub-unit(s) - used for
 PT developing prods. for studying calcium channels, e.g. for
 PT obtaining agonists and antagonists
 PS Disclosure: Page 160-162; 285pp; English.
 CC To isolate DNA encoding the Beta 1 subunit, a human hippocampus
 CC cDNA library was screened by hybridisation to a DNA fragment
 CC encoding a rabbit skeletal muscle calcium channel beta subunit.
 CC A hybridising clone was selected and was in turn used to isolate
 CC overlapping clones until the overlapping clones encompassing DNA
 CC encoding the entire human calcium channel beta 2 subunit were
 CC isolated and sequenced. Five alternatively spliced forms of the
 CC beta 1 subunit have been identified. These forms are designated
 CC beta 1-1, expressed in skeletal muscle, beta 1-2, expressed in the
 CC CNS, beta 1-3, also expressed in the CNS, beta 1-4, expressed in
 CC aorta tissue and HEK 293 cells, and beta 1-5, expressed in HEK 293
 CC cells. Full-length DNA clones encoding the beta 1-2 and beta 1-3
 CC subunits have been constructed. The subunits beta 1-1, beta 1-2,
 CC beta 1-4 and beta 1-5 have been identified by nucleic acid
 CC amplification analysis as alternatively spliced forms of the
 CC beta subunit. Sequences of the beta 1 splice variants are set
 CC forth in Q87838/R72611, Q87839/R72612 and Q87831-Q87833 and
 CC R72604-R72706.
 SQ Sequence 478 AA;
 Query Match 79.4%; Score 54; DB 14; Length 478;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 44 stsdtdtsnsvf 55
 |||:||||:|
 QY 81 STSNDTTSAAVF 92
 RESULT 9
 ID R33551 standard; Protein; 478 AA.
 AC R33551.
 DT 30-JUN-1993 (first entry)
 DE Sequence of splice variant beta 1-2 of beta human calcium
 DE channel subunit.
 KW Human calcium channel subunit; diagnosis; agonist; antagonist;
 KW Lambert Eaton syndrome.
 OS Homo sapiens.
 PN W09304083-A.
 PD 04-MAR-1993.
 PF 14-AUG-1992; U06903.
 PR 15-AUG-1991; US-745206.
 PR 10-APR-1992; US-868354.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, Mccue AF,
 PI Williams ME;
 DR WPI: 93-093836/11.
 DR N-PSDB: Q37819.
 PT DNA encoding specific human calcium channel sub-units - used for
 PT identifying calcium channel agonists and antagonists and
 PT diagnosing Lambert Eaton syndrome

PS Disclosure: Page 129-131; 150pp; English.
 CC Five alternatively spliced forms of the human calcium channel
 CC beta 1 subunit have been identified and DNA encoding a number 1-1,
 CC of forms have been isolated. These forms are designated beta 1-1,
 CC expressed in skeletal muscle, beta 1-2, expressed in the CNS, beta
 CC 1-3, also expressed in the CNS, beta 1-4, expressed in aorta tissue
 CC and HEK 293 cells, and beta 1-5, expressed in HEK 293 cells.
 SQ Sequence 478 AA;
 Query Match 79.4%; Score 54; DB 6; Length 478;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 44 stsdtdtsnsvf 55
 |||:||||:|
 QY 81 STSNDTTSAAVF 92
 RESULT 10
 ID R72604 standard; Protein; 479 AA.
 AC R72604;
 DT 01-DEC-1995 (first entry)
 DE Human calcium channel subunit beta 1-1.
 KW Calcium channel subunit; antagonist; agonist; diagnosis;
 KW Lambert Eaton Syndrome.
 OS Homo sapiens.
 PN W09504822-A.
 PD 16-FEB-1995.
 PF 11-AUG-1994; U09230.
 PR 11-AUG-1993; US-105536.
 PR 05-NOV-1993; US-149097.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Ellis SB, Gillespie A, Harpold MM, Mccue AF, Williams ME;
 DR WPI: 95-090900/12.
 DR N-PSDB: Q87831.
 PT DNA encoding human calcium channel sub-unit(s) - used for
 PT developing prods. for studying calcium channels, e.g. for
 PT obtaining agonists and antagonists
 PS Disclosure: Page 253-256; 285pp; English.
 CC To isolate DNA encoding the Beta 1 subunit, a human hippocampus
 CC cDNA library was screened by hybridisation to a DNA fragment
 CC encoding a rabbit skeletal muscle calcium channel beta subunit.
 CC A hybridising clone was selected and was in turn used to isolate
 CC overlapping clones until the overlapping clones encompassing DNA
 CC encoding the entire human calcium channel beta 2 subunit were
 CC isolated and sequenced. Five alternatively spliced forms of the
 CC beta 1 subunit have been identified. These forms are designated
 CC beta 1-1, expressed in skeletal muscle, beta 1-2, expressed in the
 CC CNS, beta 1-3, also expressed in the CNS, beta 1-4, expressed in
 CC aorta tissue and HEK 293 cells, and beta 1-5, expressed in HEK 293
 CC cells. Full-length DNA clones encoding the beta 1-2 and beta 1-3
 CC subunits have been constructed. The subunits beta 1-1, beta 1-2,
 CC beta 1-4 and beta 1-5 have been identified by nucleic acid
 CC amplification analysis as alternatively spliced forms of the
 CC beta subunit. Sequences of the beta 1 splice variants are set
 CC forth in Q87838/R72611, Q87839/R72612 and Q87831-Q87833 and
 CC R72604-R72706.
 SQ Sequence 479 AA;
 Query Match 79.4%; Score 54; DB 14; Length 479;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 44 stsdtdtsnsvf 55
 |||:||||:|
 QY 81 STSNDTTSAAVF 92
 RESULT 11
 ID W63156 standard; Protein; 523 AA.
 AC W63156;
 DT 12-OCT-1998 (first entry)
 DE Human calcium channel beta-1 subunit.

CC be alternatively spliced when transcribed, giving more than one form of
 CC the protein from the same transcript, each having slightly different
 CC properties. In addition, the reactivity of the alpha 1 subunit with IgG
 CC molecules from the serum of an individual with Lambert Eaton Syndrome
 CC (LES) can be used as a diagnostic for the disease.
 SQ Sequence 219 AA;

Query Match 79.4%; Score 54; DB 33; Length 219;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 44 stsdtsnsfv 55
 |||:||||:|
 QY 81 STSNDTSAAFV 92

RESULT 5
 ID R39697 standard; Protein; 240 AA.
 AC R39697;
 DT 19-JAN-1994 (first entry)
 DE Myasthenic antigenic polypeptide.
 KW MAP: antibody detection; LEMS; Lambert-Eaton myasthenic syndrome;
 KW paraneoplastic sensory neuropathy; p mys B; mysB; ss.
 OS Homo sapiens.
 PN WO9314098-A.
 PD 22-JUL-1993.
 PF 11-JAN-1993; U00227.
 PR 10-JAN-1992; US-820312.
 PA (SLOK) SLOAN KETTERING INST CANCER.
 PI Furneaux HM, Posner JB;
 DR WPI; 93-243126/30.
 DR N-PSDB; Q46673.
 PT New purified myasthenic antigenic polypeptide and its corresp.
 PT antibody - useful for diagnosing and treating proliferation of
 PT neoplastic cells in patient with Lambert-Eaton myasthenic syndrome
 PS Claim 5: Page 26-27; 48pp; English.
 CC The sequence is that of myasthenic antigenic polypeptide (MAP)
 CC encoded by the cDNA clone p mysB. MAP can be used to detect
 CC antibodies associated with paraneoplastic sensory neuropathy
 CC such as Lambert-Eaton myasthenic syndrome (LEMS). These antibodies
 CC are used to determine if a patient with neurological symptoms has a
 CC tumour expressing MAP, to inhibit proliferation of neoplastic cells
 CC in patients with LEMS and for imaging neoplastic cells in LEMS
 CC patients.
 SQ Sequence 240 AA;

Query Match 79.4%; Score 54; DB 8; Length 240;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 17 stsdtsnsfv 28
 |||:||||:|
 QY 81 STSNDTSAAFV 92

RESULT 6
 ID R39563 standard; Protein; 478 AA.
 AC R39563;
 DT 09-FEB-1994 (first entry)
 DE Human neuronal VDCC beta-subunit encoded by clone HB1.
 KW Voltage-dependent calcium channel; VDCC; beta-subunit;
 KW calcium flux; ss.
 OS Homo sapiens.
 PN DE422126-A.
 PD 19-AUG-1993.
 PF 06-JUL-1992; 222126.
 PR 17-FEB-1992; DE-204716.
 PR 06-JUL-1992; DE-222126.
 PA (FARB) BAYER AG.
 PI Spreyer P, Unterbeck A;
 DR WPI; 93-265734/34.
 DR N-PSDB; Q46076.
 PT Human neuronal beta-unit cDNA of voltage dependent calcium

PT channels - useful in calcium-flux studies and screening systems
 PT for agonists and antagonists of calcium channels
 PS Claim 2: Page 5-7; 13pp; German
 CC A first oligonucleotide probe (Q46075) complementary to nucleotides
 CC 361-400 of the VDCC beta-subunit from rabbit skeletal muscle was
 CC used to screen a human hippocampus cDNA library. A 1.9kb cDNA
 CC fragment was isolated for further screening to isolate human VDCC
 CC beta-subunit coding sequences. Clone HB1 was sequenced and found
 CC to have 92% homology to the rabbit beta-subunit. The amino acid
 CC sequence R39563 was deduced from the open reading frame.
 CC See also Q46077-Q46078.
 SQ Sequence 478 AA;

Query Match 79.4%; Score 54; DB 8; Length 478;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtsnsfv 55
 |||:||||:|
 QY 81 STSNDTSAAFV 92

RESULT 7
 ID W63143 standard; Protein; 478 AA.
 AC W63143;
 DT 12-OCT-1998 (first entry)
 DE Human calcium channel beta subunit splice variant beta2.
 DE Beta subunit; human; calcium channel; assay; detection;
 KW characterisation; Lambert Eaton Syndrome; LES; diagnosis.
 KW Homo sapiens.
 OS Homo sapiens.
 PN US5792846-A.
 PD 11-AUG-1998; 455543.
 PR 31-MAY-1995; US-223305.
 PR 04-APR-1994; US-176899.
 PR 04-APR-1988; US-603751.
 PR 04-APR-1989; WO-U01408.
 PR 20-FEB-1990; US-482384.
 PR 30-NOV-1990; US-620250.
 PR 15-AUG-1991; US-745206.
 PR 31-MAY-1995; US-455543.
 PA (SIBI-) SIBIA NEUROSCIENCES INC.
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
 PI Williams ME;
 DR WPI; 98-456192/39.
 DR N-PSDB; V42687.

PT DNA encoding human calcium channel alpha 1B subunit protein -
 PT useful for recombinant production of the channel for screening of
 PT its modulators, and diagnosis of Lambert Eaton Syndrome
 PS Claim 3: Columns 217-220; 166pp; English.
 CC The present sequence is encoded by a splice variant of the beta subunit
 CC of a human calcium channel. Calcium channels are membrane-spanning,
 CC multi-subunit proteins that allow controlled entry of calcium ions into
 CC cells. This leads to depolarisation events required for muscle
 CC contraction. The recombinant subunit, when expressed with nucleic acids
 CC encoding the complete calcium channel, can be used in assays for the
 CC detection and characterisation of compounds that modulate the channel.
 CC The DNA encoding the subunits can be alternatively spliced when
 CC transcribed, giving more than one form of the protein from the same
 CC transcript, each having slightly different properties. In addition, the
 CC reactivity of the alpha 1 subunit with IgG molecules from the serum of
 CC an individual with Lambert Eaton Syndrome (LES) can be used as a
 CC diagnostic for the disease.
 SQ Sequence 478 AA;

Query Match 79.4%; Score 54; DB 33; Length 478;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtsnsfv 55
 |||:||||:|
 QY 81 STSNDTSAAFV 92

```

SQ Sequence 216 AA:
Query Match 79.4%; Score 54; DB 33; Length 216;
Best Local Similarity 75.0%; Pred. No. 1.78e+01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtsnsfv 55
   |||:||||:|
QY 81 STSNDTSAAFV 92

RESULT 2
ID R72605 standard; Protein; 216 AA.
AC R72605;
DE 01-DEC-1995 (first entry)
DE Human calcium channel subunit beta 1-4.
KW Calcium channel subunit; antagonist; agonist; diagnosis;
KW Lambert Eaton Syndrome.
OS Homo sapiens.
PN W09504822-A.
PD 16-FEB-1995.
PF 11-AUG-1994; U09230.
PR 11-AUG-1993; US-105536.
PR 05-NOV-1993; US-149097.
PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.
PI Ellis SB, Gillespie A, Harpold MM, McCue AF, Williams ME;
DR WPI: 95-090900/12.
DR N-PSDB; Q87833.
DR 44 stsdtsnsfv 55
   |||:||||:|
QY 81 STSNDTSAAFV 92

Query Match 79.4%; Score 54; DB 14; Length 219;
Best Local Similarity 75.0%; Pred. No. 1.78e+01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtsnsfv 55
   |||:||||:|
QY 81 STSNDTSAAFV 92

RESULT 4
ID W63158 standard; Protein; 219 AA.
AC W63158;
DE 12-OCT-1998 (first entry)
DE Human calcium channel beta-5 subunit.
DE Beta-5 subunit; human; calcium channel; assay; detection;
KW Characterisation; Lambert Eaton Syndrome; LES; diagnosis; ds.
KW Homo sapiens.
PN U55792846-A.
PD 11-AUG-1998.
PF 31-MAY-1995; 455543.
PR 04-APR-1994; US-223305.
PR 04-APR-1988; US-176899.
PR 04-APR-1989; US-603751.
PR 04-APR-1989; US-603751.
PR 20-FEB-1990; WO-001408.
PR 30-NOV-1990; US-620250.
PR 15-AUG-1991; US-745206.
PR 31-MAY-1995; US-455543.
PA (SIBI-) SIBIA NEUROSCIENCES INC.
PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
PI Williams ME;
DR WPI: 98-456192/39.
DR N-PSDB; V42707.
DR DNA encoding human calcium channel alpha 1B subunit protein -
PT useful for recombinant production of the channel for screening of
PT its modulators, and diagnosis of Lambert Eaton Syndrome
PS Disclosure: Columns 227-230; 166pp; English.
PS The present sequence represents the beta-5 subunit of a human calcium
CC channel. Calcium channels are membrane-spanning, multi-subunit proteins
CC that allow controlled entry of calcium ions into cells. This leads
CC to depolarisation events required for muscle contraction. The recombinant
CC subunit, when expressed with nucleic acids encoding the complete calcium
CC channel, can be used in assays for the detection and characterisation of
CC compounds that modulate the channel. The DNA encoding the subunits can

```

 W P S R E H

 (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:33:07 1999; MasPar time 4.43 Seconds

Tabular output not generated. 62.340 Million cell updates/sec

Title: >PCT-US99-13024-2

Description: (81-93) from PCTUS9913024 pep (12 of 12)

Sequence: 1 STSNDTTSAAFPVS 13

Scoring table: PAM 150
 Gap 11

Searched: 170751 seqs, 2126608 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-gensseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33
 34:part34 35:part35 36:part36 37:part37 38:part38
 39:part39

Statistics: Mean 15.573; Variance 46.313; scale 0.336

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	54	79.4	216	33	W63157 Human calcium channel	1.78e+01
2	54	79.4	216	14	R72605 Human calcium channel	1.78e+01
3	54	79.4	219	14	R72606 Human calcium channel	1.78e+01
4	54	79.4	219	33	W63158 Human calcium channel	1.78e+01
5	54	79.4	240	8	R39697 Myasthenic antigenic	1.78e+01
6	54	79.4	478	8	R39563 Human neuronal VDCC b	1.78e+01
7	54	79.4	478	33	W63143 Human calcium channel	1.78e+01
8	54	79.4	478	14	R72611 Human calcium channel	1.78e+01
9	54	79.4	478	6	R32551 Sequence of splice va	1.78e+01
10	54	79.4	479	14	R72604 Human calcium channel	1.78e+01
11	54	79.4	523	33	W63156 Human calcium channel	1.78e+01
12	54	79.4	530	33	W63147 Human calcium channel	1.78e+01
13	54	79.4	571	8	R39565 Human neuronal VDCC b	1.78e+01
14	54	79.4	598	33	W63144 Human calcium channel	1.78e+01
15	54	79.4	598	14	R72612 Human calcium channel	1.78e+01
16	54	79.4	598	6	R33552 Sequence of splice va	1.78e+01

17	54	79.4	660	14	R72613	Human neuronal calcium	1.78e+01
18	44	64.7	636	36	W79589	Human Kv potassium ch	2.29e+02
19	44	64.7	636	36	W79591	Human Kv potassium ch	2.29e+02
20	44	64.7	655	36	W79590	Human Kv potassium ch	2.29e+02
21	44	64.7	1130	15	R71729	Merosin major subunit	2.29e+02
22	44	64.7	1130	3	R13436	Merosin M polypeptide	2.29e+02
23	44	64.7	3110	15	R71730	Merosin major subunit	2.29e+02
24	43	63.2	757	20	W03179	Bovine poly-immunoglob	2.93e+02
25	42	61.8	127	28	W31716	Gamma-1 heavy chain a	3.74e+02
26	42	61.8	653	21	W14264	2. japonica phosphoen	3.74e+02
27	42	61.8	953	19	R97284	Human 26S proteasome	3.74e+02
28	42	61.8	2783	5	R23963	AFP-1 (Ala 2460 Val)	3.74e+02
29	42	61.8	2783	5	R23962	AFP-1.	3.74e+02
30	41	60.3	98	13	R72068	DP10 VH region.	4.76e+02
31	41	60.3	120	25	W27550	Human Ab heavy chain	4.76e+02
32	41	60.3	120	10	R54795	SPA-reactive IgM heav	4.76e+02
33	41	60.3	123	24	W19881	CEA-specific antibody	4.76e+02
34	41	60.3	123	24	W19889	CEA-specific antibody	4.76e+02
35	41	60.3	123	24	W19888	CEA-specific antibody	4.76e+02
36	41	60.3	123	24	W19887	CEA-specific antibody	4.76e+02
37	41	60.3	128	28	W31715	Gamma-1 heavy chain a	4.76e+02
38	41	60.3	268	27	W27136	Achromobacter lyticus	4.76e+02
39	41	60.3	268	27	W31403	Achromobacter lyticus	4.76e+02
40	41	60.3	268	26	W27135	Achromobacter lyticus	4.76e+02
41	41	60.3	397	19	W04270	B.t. alkaline proteas	4.76e+02
42	41	60.3	481	5	R24442	Sequence of antibody	4.76e+02
43	41	60.3	512	39	W87797	Protease encoded by c	4.76e+02
44	41	60.3	512	14	R80505	S. lividans protease	4.76e+02
45	41	60.3	1170	12	R63231	Crystal protein CrYT	4.76e+02

ALIGNMENTS

RESULT 1
 ID W63157 standard; Protein; 216 AA.

DE 12-OCT-1998 (first entry)
 DT Human calcium channel beta-4 subunit.
 KW Beta-4 subunit; human; calcium channel; assay; detection;
 KW Characterisation; Lambert Eaton Syndrome; LES; diagnosis.
 OS Homo sapiens.

PN US5792846-A.
 PD 11-AUG-1998.

PF 31-MAY-1995; 455543.

PR 04-APR-1994; US-223305.

PR 04-APR-1988; US-176899.

PR 04-APR-1989; US-603751.

PR 04-APR-1989; WO-U01408.

PR 20-FEB-1990; US-482384.

PR 30-NOV-1990; US-620250.

PR 15-AUG-1991; US-745206.

PR 31-MAY-1995; US-455543.

PA (SIBI-) SIBIA NEUROSCIENCES INC.

PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,

PI Williams ME;

DR WPI: 98-456192/39.

DR N-PSDB: V43706.

PT DNA encoding human calcium channel alpha 1B sub-unit protein -
 PT useful for recombinant production of the channel for screening of
 PT its modulators, and diagnosis of Lambert Eaton Syndrome
 PS Disclosure: Columns 225-228; 166pp; English.

CC The present sequence represents the beta-4 subunit of a human calcium
 CC channel. Calcium channels are membrane-spanning, multi-subunit proteins
 CC that allow controlled entry of calcium ions into cells. This leads
 CC to depolarisation events required for muscle contraction. The recombinant
 CC subunit, when expressed with nucleic acids encoding the complete calcium
 CC channel, can be used in assays for the detection and characterisation of
 CC compounds that modulate the channel. The DNA encoding the subunits can
 CC be alternatively spliced when transcribed, giving more than one form of
 CC the protein from the same transcript, each having slightly different
 CC properties. In addition, the reactivity of the alpha 1 subunit with IgG
 CC molecules from the serum of an individual with Lambert Eaton Syndrome
 CC (LES) can be used as a diagnostic for the disease.

Query Match 58.1%; Score 50; DB 2; Length 371;
Best Local Similarity 53.8%; Pred. No. 9.22e+00;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 164 DGTGNLIQSSSY 176
QY 66 DPSTGALVDSKSY 78

RESULT 14
ID O70461 PRELIMINARY; PRT; 492 AA.
AC O70461;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE MONOCARBOXYLATE TRANSPORTER MCT3.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-RETINA;
RA PHILP N.J., YOON H.;
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF059258; G3170609;
SQ SEQUENCE 492 AA; 51590 MW; 20F051C1 CRC32;

Query Match 58.1%; Score 50; DB 11; Length 492;
Best Local Similarity 61.5%; Pred. No. 9.22e+00;
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 373 PSAGRLVDALKNY 385
QY 67 PSTGALVDS-KSY 78

RESULT 15
ID O13151 PRELIMINARY; PRT; 542 AA.
AC O13151;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE MONOCARBOXYLATE TRANSPORTER 3.
GN MCT3.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LEGHORN;
RA PHILP N.J., YOON H.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF000240; G2198807;
SQ SEQUENCE 542 AA; 58085 MW; 836DA7B1 CRC32;

Query Match 58.1%; Score 50; DB 13; Length 542;
Best Local Similarity 61.5%; Pred. No. 9.22e+00;
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 410 PSAGRLVDALKNY 422
QY 67 PSTGALVDS-KSY 78

Search completed: Wed Sep 1 16:29:45 1999
Job time : 21 secs.


```

QY 66 DPSTGALVDSKSYA 79
      || |||::| |::|
RESULT 10 PRELIMINARY; PRT; 142 AA.
ID Q87600;
AC Q87600;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS SIMIAN IMMUNODEFICIENCY VIRUS (SIV-AGM).
OS VIRUSES; RETROD VIRUSES; RETROVIRIDAE; LENTIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-P081;
RA BIBOLLET-RUCHE F., BENGUES C., GALAT-LUONG A., GALAT G., POURRUT X.,
RA VIDAL N., VEAS F., DURAND J.P., CUNY G.;
RL J. VIROL. 71:307-313(1997).
DR EMBL: U37209; G1353456; -.
DR PFAM: PF00517; GP41; 1.
KW ENVELOPE PROTEIN.
FT NON_TER 1
FT NON_TER 142
SQ SEQUENCE 142 AA; 16852 MW; C58A9A91 CRC32;

Query Match 58.1%; Score 50; DB 14; Length 142;
Best Local Similarity 46.7%; Pred. No. 9.22e+00;
Matches 7; Conservative 5; Mismatches 2; Indels 1; Gaps 1;

Db 86 EPDSGGLSRDSRSY 100
      :|::| |::|
QY 66 DPSTGALV-DSKSYA 79

RESULT 11 PRELIMINARY; PRT; 156 AA.
ID Q35308;
AC Q35308;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE MONOCARBOXYLATE TRANSPORTER MCT3 (FRAGMENT).
OS MUS MUSCULUS (MOUSE).
OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OS SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C3H;
RA PHILP N.J., YOON H.;
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF019111; G2407664; -.
FT NON_TER 1
SQ SEQUENCE 156 AA; 16253 MW; DB394B9E CRC32;

Query Match 58.1%; Score 50; DB 11; Length 156;
Best Local Similarity 61.5%; Pred. No. 9.22e+00;
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 37 PSAGRLVDALKNY 49
      |||::| |::|
QY 67 PSTGALVDS-KSY 78

RESULT 12 PRELIMINARY; PRT; 262 AA.
ID Q44188;
AC Q44188;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE W-AMINO-TRANSFERASE-LIKE PROTEIN.
GN OATA.
OS AGROBACTERIUM RADIOBACTER.

```

```

OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; AGROBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T1305 LAC9;
RX MEDLINE; 96425889.
RA TIBURTIUS A., DE LUCA N., HUSSAIN H., JOHNSTON A.W.B.;
RT "Expression of the exy gene, required for exopolysaccharide
RT synthesis in Agrobacterium, is activated by the regulatory ros
RT gene.";
RL MICROBIOLOGY 142:2621-2629(1996).
DR EMBL: X95394; E220383; -.
DR PFAM: PF00202; aminotran_3; 1.
KW TRANSFERASE.
SQ SEQUENCE 262 AA; 28124 MW; DF7A4568 CRC32;

Query Match 58.1%; Score 50; DB 2; Length 262;
Best Local Similarity 63.6%; Pred. No. 9.22e+00;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 110 STGVLVPPKGY 120
      |||::| |::|
QY 68 STGALVDSKSY 78

RESULT 13 PRELIMINARY; PRT; 371 AA.
ID P72097;
AC P72097; P72099; P72100; P72101;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CMP-N-ACETYLNEURAMINATE-BETA-GALACTOSAMIDE-ALPHA-2,3-
DE SIALYLTRANSFERASE (EC 2.4.99.-)
DE (BETA-GALACTOSIDE ALPHA-2,3-SIALYLTRANSFERASE) (ALPHA 2,3-ST)
DE (LIPOOLIGOSACCHARIDE SIALYLTRANSFERASE) (LST).
GN LST.
OS NEISSERIA MENINGITIDIS.
OC BACTERIA; PROTEOBACTERIA; BETA SUBDIVISION; NEISSERIACEAE; NEISSERIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MC58 / NRCC 4728, 406Y / NRCC 4030, AND M982B / NRCC 4725;
RA GILBERT M., WATSON D.C., CUNNINGHAM A.-M., JENNINGS M.P., YOUNG N.M.,
RA WAKARCHUK W.W.;
RL J. BIOL. CHEM. 271:28271-28276(1996).
CC -1- FUNCTION: TRANSFERS SIALIC ACID FROM THE SUBSTRATE CMP-SIALIC
CC ACID DONOR TO THE TERMINAL
CC BETA-D-GALACTOSYL-1,4-ACETYL-BETA-D-GLUCOSAMINE ON THE
CC LACTO-N-NEOTETRAOSE BRANCH OF THE LIPOOLIGOSACCHARIDE.
CC -1- CATALYTIC ACTIVITY: CMP-N-ACETYLNEURAMINATE +
CC BETA-D-GALACTOSYL-1,4-ACETYL-BETA-D-GLUCOSAMINE -> CMP +
CC ALPHA-N-ACETYLNEURAMINYL-2,3-BETA-D-GALACTOSYL-1,4-N-ACETYL-BETA-
CC D-GLUCOSAMINE.
CC -1- PATHWAY: GLYCOSYLATION.
CC EMBL: U60660; G1546004; -.
DR EMBL: U60661; G1546007; -.
DR EMBL: U60662; G1546009; -.
DR EMBL: U60663; G1546011; -.
KW TRANSFERASE; GLYCOSYLTRANSFERASE.
FT VARIANT 2 2 G -> S (IN STRAIN M982B / NRCC 4725).
FT VARIANT 29 29 Q -> H
FT (IN STRAINS 406Y / NRCC 4030 AND M982B /
FT NRCC 4725).
FT E -> D
FT (IN STRAINS 406Y / NRCC 4030 AND M982B /
FT NRCC 4725).
FT N -> K
FT (IN STRAINS 406Y / NRCC 4030 AND M982B /
FT NRCC 4725).
FT L -> *; (IN STRAIN M982B / NRCC 4725;
FT LOSS OF ACTIVITY).
FT T -> A (IN STRAIN 406Y / NRCC 4030).
FT K -> N (IN STRAIN 406Y / NRCC 4030).
SQ SEQUENCE 371 AA; 42611 MW; 73DB83A9 CRC32;

```

```

RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 98290545.
RA NAGASE T., ISHIKAWA K., MIYAJIMA N., TANAKA A., KOTANI H., NOMURA N.,
RA OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RL DNA RES. 5:31-39(1998).
DR EMBL; AB011135; D1026419; -.
SQ SEQUENCE 870 AA; 94665 MW; EB9CD3F4 CRC32;

Query Match 60.5%; Score 52; DB 4; Length 870;
Best Local Similarity 42.9%; Pred. No. 3.35e+00;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 798 EPAQDLSVQSESYT 811
   |||:||||:
QY 66 DPSTGALVDSKSYA 79

RESULT 7
ID Q24281 PRELIMINARY; PRT; 540 AA.
AC Q24281;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PHOTOLYASE.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PHR-; TISSUE=OVARY;
RX MEDLINE; 96178677.
RA TODO T., RYO H., YAMAMOTO K., TOH H., INUI T., AYAKI H., NOMURA T.,
RA IKENAGA M.;
RT "Similarity among the Drosophila (6-4)photolyase, a human photolyase
RT homolog, and the DNA photolyase-blue-light photoreceptor family."
RL SCIENCE 272:109-112(1996).
DR EMBL; D83701; D1012739; -.
KW LYASE.
SQ SEQUENCE 540 AA; 62548 MW; 62F139EE CRC32;

Query Match 59.3%; Score 51; DB 5; Length 540;
Best Local Similarity 35.7%; Pred. No. 5.58e+00;
Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 456 EPKASLVQDORAYG 469
   .|||:||||:
QY 66 DPSTGALVDSKSYA 79

RESULT 8
ID Q88274 PRELIMINARY; PRT; 1340 AA.
AC Q88274;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE POLYPROTEIN (FRAGMENT).
OS SWEET POTATO FEATHERY MOTTLE VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
OC POTYVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEVERE;
RX MEDLINE; 95390803.
RA MORI M., SAKAI J., KIMURA T., USUGI T., HAYASHI T., HANADA K.,
RA NISHIGUCHI M.;
RT "Nucleotide sequence analysis of two nuclear inclusion body and coat
RT protein genes of a sweet potato feathery mottle virus severe strain
RT (SPFMV-S) genomic RNA."

```

```

RL ARCH. VIROL. 140:1473-1482(1995).
DR EMBL; D38543; D1008127; -.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00767; Poly_coat; 1.
DR PFAM; PF00863; Peptidase_C4; 1.
KW POLYPROTEIN; COAT PROTEIN.
FT NON_TER 1
FT CHAIN 1 16 CYTOPLASMIC INCLUSION PROTEIN.
FT CHAIN 17 69 6K2 PROTEIN.
FT CHAIN 70 261 VPG PROTEIN.
FT CHAIN 262 504 NIA PROTEASE.
FT CHAIN 505 1025 NUCLEAR INCLUSION PROTEIN B.
FT CHAIN 1026 1340 COAT PROTEIN.
SQ SEQUENCE 1340 AA; 151873 MW; 12C01867 CRC32;

Query Match 59.3%; Score 51; DB 14; Length 1340;
Best Local Similarity 50.0%; Pred. No. 5.58e+00;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 146 DPLTGAVIDDSPYT 159
   |||:||||:
QY 66 DPSTGALVDSKSYA 79

RESULT 9
ID O39734 PRELIMINARY; PRT; 3493 AA.
AC O39734;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POLYPROTEIN.
OS SWEET POTATO FEATHERY MOTTLE VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
OC POTYVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S STRAIN;
RA NISHIGUCHI M.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S STRAIN;
RX MEDLINE; 98336489.
RA SAKAI J., MORI M., MORISHITA T., TANAKA M., HANADA K., USUGI T.,
RA NISHIGUCHI M.;
RT "Complete nucleotide sequence and genome organization of sweet potato
RT feathery mottle virus (S strain) genomic RNA: the large coding region
RT of the P1 gene."
RL ARCH. VIROL. 142:1553-1562(1997).
DR EMBL; D86371; D1023570; -.
DR PFAM; PF00271; helicase_C; 1.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00767; Poly_coat; 1.
DR PFAM; PF00851; Peptidase_C6; 1.
DR PFAM; PF00863; Peptidase_C4; 1.
KW POLYPROTEIN.
FT CHAIN 1 664 POTENTIAL.
FT CHAIN 665 1122 POTENTIAL.
FT CHAIN 1123 1474 POTENTIAL.
FT CHAIN 1475 1526 POTENTIAL.
FT CHAIN 1527 2169 POTENTIAL.
FT CHAIN 2170 2222 POTENTIAL.
FT CHAIN 2223 2414 GENOME-LINKED VIRAL PROTEIN.
FT CHAIN 2415 2657 PROTEASE.
FT CHAIN 2658 3178 REPLICASE.
FT CHAIN 3179 3493 COAT PROTEIN.
SQ SEQUENCE 3493 AA; 393818 MW; A02EFA05 CRC32;

Query Match 59.3%; Score 51; DB 14; Length 3493;
Best Local Similarity 50.0%; Pred. No. 5.58e+00;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 2299 DPLTGAVIDDSPYT 2312

```

ID Q46094 PRELIMINARY; PRT; 368 AA.
AC Q46094;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE K1AA0574 PROTEIN (FRAGMENT).
DE HEAT SHOCK PROTEIN/SERINE PROTEASE (FRAGMENT).
GN HTRA.
OS CAMPYLOBACTER JEJUNI.
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; CAMPYLOBACTER GROUP;
OC CAMPYLOBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UA580;
RX MEDLINE; 90384493.
RA TAYLOR D.E., HIRATSUKA K.;
RT "Use of non-radioactive DNA probes for detection of Campylobacter
jejunii and Campylobacter coli in stool specimens.";
RL MOL. CELL. PROBES 4:261-271(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-UA580;
RA HIRATSUKA K.;
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U27271; G881375; -;
DR PFAM; PF00089; trypsin; 1.
DR PFAM; PF00595; PDZ; 2.
KW HEAT SHOCK; PROTEASE.
FT NON_TER 1
SQ SEQUENCE 368 AA; 39491 MW; A5062589 CRC32;

Query Match 65.1%; Score 56; DB 2; Length 368;
Best Local Similarity 72.7%; Pred. No. 4.07e-01;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 121 SGGALVDSRGY 131
I :|||||:
QY 68 STGALVDSKSY 78

RESULT 3
ID Q46120 PRELIMINARY; PRT; 472 AA.
AC Q46120;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SERINE PROTEASE.
GN HTRA.
OS CAMPYLOBACTER JEJUNI.
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; CAMPYLOBACTER GROUP;
OC CAMPYLOBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-81116;
RA HENDERSON J., WOOD A.C., EMERY M.J., WREN B.W., KETLEY J.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X82628; E315368; -;
DR PFAM; PF00089; trypsin; 1.
DR PFAM; PF00595; PDZ; 2.
KW PROTEASE.
SQ SEQUENCE 472 AA; 50940 MW; 6CBCA101 CRC32;

Query Match 65.1%; Score 56; DB 2; Length 472;
Best Local Similarity 72.7%; Pred. No. 4.07e-01;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 225 SGGALVDSRGY 235
I :|||||:
QY 68 STGALVDSKSY 78

RESULT 4
ID O60320 PRELIMINARY; PRT; 405 AA.
AC O60320;

DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE K1AA0574 PROTEIN (FRAGMENT).
GN K1AA0574.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN.
RX MEDLINE; 98290545.
RA NAGASE T., ISHIKAWA K., MIYAJIMA N., TANAKA A., KOTANI H., NOMURA N.,
RA OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";
RL DNA RES. 5:31-39(1998).
DR EMBL; AB011146; D1026430; -;
FT NON_TER 1
SQ SEQUENCE 405 AA; 42409 MW; 09BA9932 CRC32;

Query Match 60.5%; Score 52; DB 4; Length 405;
Best Local Similarity 77.8%; Pred. No. 3.35e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 345 SLVDSKAYA 353
:|||||:
QY 71 ALVDSKSYA 79

RESULT 5
ID O46178 PRELIMINARY; PRT; 552 AA.
AC O46178;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE RADIAL SPOKEHEAD.
OS STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; EUECHINOIDEA;
OC ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE; STRONGYLOCENTROTUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98119758.
RA GINGRAS D., GAGNON C.;
RT "Molecular cloning and characterization of a radial spoke head
protein of sea urchin sperm axonemes: involvement of the protein in
the regulation of sperm motility.";
RL MOL. BIOL. CELL 9:513-522(1998).
DR EMBL; U73123; G2905895; -;
SQ SEQUENCE 552 AA; 62723 MW; 898CFCCC CRC32;

Query Match 60.5%; Score 52; DB 5; Length 552;
Best Local Similarity 50.0%; Pred. No. 3.35e+00;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 39 PTEALVNAKAY 50
I :|||||:
QY 67 PSTGALVDSKSY 78

RESULT 6
ID O60309 PRELIMINARY; PRT; 870 AA.
AC O60309;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE K1AA0563 PROTEIN.
GN K1AA0563.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]

W A I S R A H

(TW)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_Pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:29:24 1999; MasPar time 5.26 Seconds
Tabular output not generated. 145.242 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (66-79) from PCTUS9913024.pep (11 of 12)
Perfect Score: 86
Sequence: 1 DPSTGALVDSKSYA 14

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sprembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 24.488; Variance 26.991; scale 0.907

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	59	68.6	1577	2 Q54178	GLUCOSYLTRANSFERASE	7.85e-02
2	56	65.1	368	2 Q46094	HEAT SHOCK PROTEIN/SER	4.07e-01
3	56	65.1	472	2 Q46120	SERINE PROTEASE	4.07e-01
4	52	60.5	405	4 O60320	KIAA0574 PROTEIN (FRAG	3.35e-00
5	52	60.5	552	5 O46178	RADIAL SPOKEHEAD	3.35e-00
6	52	60.5	870	4 O60309	KIAA0563 PROTEIN	3.35e-00
7	51	59.3	540	5 O24281	PHOTOLYASE	5.58e-00
8	51	59.3	1340	14 Q82874	POLYPROTEIN (FRAGMENT)	5.58e-00
9	51	59.3	3493	14 Q39734	POLYPROTEIN	5.58e-00
10	50	58.1	142	14 Q87600	ENVELOPE GLYCOPROTEIN	9.22e+00
11	50	58.1	156	11 Q35308	MONOCARBOXYLATE TRANSP	9.22e+00
12	50	58.1	262	2 Q44188	W-AMINO-TRANSFERASE-LI	9.22e+00
13	50	58.1	371	2 P72097	CMP-N-ACETYLNEURAMINAT	9.22e+00
14	50	58.1	492	11 O70461	MONOCARBOXYLATE TRANSP	9.22e+00
15	50	58.1	542	13 Q3151	RETINAL EPITHELIAL TRANSP	9.22e+00
16	50	58.1	542	13 Q30632	RETINAL EPITHELIAL MEM	9.22e+00
17	50	58.1	1133	10 O04626	SIMILARITY TO MYOSIN H	9.22e+00
18	50	58.1	1378	11 Q61138	PATERNALLY EXPRESSED P	9.22e+00
19	49	57.0	319	2 Q34966	YCDH	1.51e+01
20	49	57.0	327	14 Q92324	HEMAGGLUTININ (FRAGMEN	1.51e+01

21	49	57.0	329	14	O40846	HEMAGGLUTININ (FRAGMEN	1.51e+01
22	49	57.0	329	14	O40849	HEMAGGLUTININ (FRAGMEN	1.51e+01
23	49	57.0	427	5	Q25991	PARASITOPHOROUS VACUOL	1.51e+01
24	49	57.0	734	11	O88970	INSULIN RECEPTOR SUBST	1.51e+01
25	49	57.0	966	1	O26770	HYPOTHETICAL 104.9 KD	1.51e+01
26	49	57.0	984	5	Q26016	SERINE RICH PROTEIN (S	1.51e+01
27	49	57.0	1139	2	Q54073	ANCHOR PROTEIN, LCM	1.51e+01
28	48	55.8	195	2	O06554	HYPOTHETICAL 20.8 KD P	2.46e+01
29	48	55.8	328	14	O82581	HAEMAGGLUTININ (HA1 DO	2.46e+01
30	48	55.8	329	14	O40888	HEMAGGLUTININ (FRAGMEN	2.46e+01
31	48	55.8	329	14	O40887	HEMAGGLUTININ (FRAGMEN	2.46e+01
32	48	55.8	329	14	O70688	HEMAGGLUTININ (FRAGMEN	2.46e+01
33	48	55.8	329	14	O40901	HEMAGGLUTININ (FRAGMEN	2.46e+01
34	48	55.8	329	14	O70682	HEMAGGLUTININ (FRAGMEN	2.46e+01
35	48	55.8	329	14	O70684	HEMAGGLUTININ (FRAGMEN	2.46e+01
36	48	55.8	427	10	Q03349	S-LOCUS-SPECIFIC GLYCO	2.46e+01
37	48	55.8	477	11	O70142	SCK, PARTIAL CDS (FRAG	2.46e+01
38	48	55.8	480	2	O86047	PUTATIVE TRANSCRIPTION	2.46e+01
39	48	55.8	540	4	O60230	SCK, PARTIAL CDS (FRAG	2.46e+01
40	48	55.8	550	14	O82753	HAEMAGGLUTININ PRECURS	2.46e+01
41	48	55.8	580	3	O13902	DIHYDROXYACETONE KINAS	2.46e+01
42	48	55.8	712	2	O46520	OMPID PRECURSOR	2.46e+01
43	48	55.8	1571	11	O54978	ZINC FINGER PROTEIN	2.46e+01
44	47	54.7	329	14	O40654	HEMAGGLUTININ (FRAGMEN	3.97e+01
45	47	54.7	329	14	Q67138	HEMAGGLUTININ (FRAGMEN	3.97e+01

ALIGNMENTS

RESULT 1
ID Q54178 PRELIMINARY; PRT: 1577 AA.
AC Q54178; Q54247;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE.
GN GTFG.
OS STREPTOCOCCUS GORDONII.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CHALLIS;
RX MEDLINE; 96157084.
RA VICKERMAN M.M., SULAVIK M.C., CLEWELL D.B.;
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase phase variants.";
RL DEV. BIOL. STAND. 85:309-314(1995).
RN [2]
RP SEQUENCE OF 1-96 FROM N.A.
RC STRAIN-CHALLIS;
RX MEDLINE; 92276337.
RA SULAVIK M.C., TARDIF G., CLEWELL D.B.;
RT "Identification of a gene, rgg, which regulates expression of glucosyltransferase and influences the Spp phenotype of Streptococcus gordonii Challis.";
RL J. BACTERIOL. 174:3577-3586(1992).
DR EMBL; U12643; G1054877; -;
DR EMBL; M89776; G153795; -;
DR PFAM; PF00128; alpha-amylase; 1.
KW TRANSFERASE.
SQ SEQUENCE 1577 AA; 177805 MW; 7D58FEC2 CRC32;
Query Match 68.6%; Score 59; DB 2; Length 1577;
Best Local Similarity 69.2%; Pred. No. 7.85e-02;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 260 DAETGALVDSNEY 272

QY 66 DPSTGALVDSKSY 78

RESULT 2

[illegible]

DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE METHYL TRANSFERASE.
OS HEPATITIS E VIRUS (HEV).
OC VIRUSES: SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALICIVIRIDAE;
RN CALICIVIRUS.
CN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HEV037;
RA DONATI M.C., FAGAN E.A., HARRISON T.J.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: X98292; E1175751; -
KW TRANSFERASE.
SQ SEQUENCE 1693 AA; 28774286 CRC32;
Query Match 15.6%; Score 86; DB 14; Length 1693;
Best Local Similarity 25.4%; Pred. No. 6.27e+01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;
Db 572 FTSFVDCGAVLTNGPERHNLSEFSDASQSTMAAGPSLSLYAASAGLEVRYVAAGLDHRV 631
QY 8 FQGVQVQTFPLESSESVRYKISAGSCPLSTAGPSYVKFQDNPVGSQT-F-SAGLHLR-V 64
Db 632 PAPGVSP 638
QY 65 FDPSTGA 71
RESULT 10
ID Q45743 PRELIMINARY; PRT; 622 AA.
AC Q45743;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CRYIIC DELTA-ENDOTOXIN.
OS BACILLUS THURINGIENSIS.
OC PLASMID.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
CN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SHANGHAI 1 (S-1);
RA WU D., CAO X.L., BAI Y.Y., ARONSON A.I.;
RL FEMS MICROBIOL. LETT. 81:31-36(1991).
DR EMBL: X57252; G40286; -
DR PFAM: PF00555; endotoxin; 1.
KW PLASMID.
SQ SEQUENCE 622 AA; 69729 MW; 65E72171 CRC32;
Query Match 15.2%; Score 84; DB 2; Length 622;
Best Local Similarity 18.2%; Pred. No. 1.21e+00;
Matches 12; Conservative 26; Mismatches 26; Indels 2; Gaps 2;
Db 501 FISERYGNQGSRLPELSNPTARYTLRGNSYNLYLRVSSIGSSTIRVTINGRVYTANV 560
QY 12 YVQIPFLESSESVRYKISAGS-CPLSTAGPSYVKF-QDNPVGSQTF-SAGLHLR-VFDPST 69
Db 561 NNTTNN 566
QY 70 GALVDS 75
RESULT 11
ID P90770 PRELIMINARY; PRT; 921 AA.
AC P90770;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE C34B7.2 PROTEIN.
CN C34B7.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]

RP SEQUENCE FROM N.A.
RA HARRIS B.;
RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RX SEQUENCE FROM N.A.
RM MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.*;
RL NATURE 368:32-38(1994).
DR EMBL: Z83220; E1344511; -
SQ SEQUENCE 921 AA; 106698 MW; 21DB4C35 CRC32;
Query Match 15.2%; Score 84; DB 5; Length 921;
Best Local Similarity 24.7%; Pred. No. 1.21e+00;
Matches 19; Conservative 22; Mismatches 32; Indels 4; Gaps 4;
Db 212 ERWFVEIVHGVYQREYIEFLPIG-RISLTIIGRRSTKYAGTRFLKRGANPTGNVANYVETE 270
QY 2 EKFMREFGGYVQTFPLESSESVRYKISAGSCPLSTAGPSYVKFQDNPVGS-QTFSAGL 60
Db 271 Q1-VYDMASSGNVADGR 286
QY 61 HLRVFD-PSTGALVDSK 76
RESULT 12
ID Q49984 PRELIMINARY; PRT; 180 AA.
AC Q49984;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE PUTATIVE ETHYLENE RECEPTOR (FRAGMENT).
CN BOERS.
OS BRASSICA OLERACEA (CAULIFLOWER).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; BRASSICA.
CN [1]
RP SEQUENCE FROM N.A.
RA CHARNG Y.Y., SUN C.W., YAN S.L., CHOU S.J., CHEN Y.R., YANG S.F.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U87239; G2738025; -
FT NONTER 1
FT NONTER 180
SQ SEQUENCE 180 AA; 19859 MW; D9B072C1 CRC32;
Query Match 15.1%; Score 83; DB 10; Length 180;
Best Local Similarity 36.0%; Pred. No. 1.68e+00;
Matches 18; Conservative 9; Mismatches 20; Indels 3; Gaps 3;
Db 48 IPHTCPLAKIGPPVGVKFAPEVSVRVPL-LHLSNFGSDMSDL-SGRKY 95
QY 30 IAGSCPLSTAGPSYVKFQDNPVGSQTF-SAGLHLR-VFDPST-GALVDSKSY 78
RESULT 13
ID Q38022 PRELIMINARY; PRT; 281 AA.
AC Q38022;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ORF 1.
OS BACTERIOPHAGE PHI-C31.

[illegible]

WATERMAN
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Sep 1 16:13:05 1999; MasPar time 9.45 seconds
Tabular output not generated. 450.572 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (1-78) from PCTUS9913024.pep (5 of 12)
Sequence: 1 MEKFAEFGQGYVTPFLSE.....GLHLRVDPSTGALVDSKY 78

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 37.051; Variance 67.149; scale 0.552

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	100	18.1	1693	14	COMPLETE GENOME SEQUEN	4.78e-03
2	96	17.4	114	14	A606L PROTEIN.	2.00e-02
3	92	16.7	299	14	HYPOTHETICAL 32.4 KD P	8.15e-02
4	92	16.7	300	14	UL7 POLYPEPTIDE.	8.15e-02
5	89	16.2	566	2	HYPOTHETICAL 63.5 KD P	2.28e-01
6	86	15.6	1693	14	ORF 1.	6.27e-01
7	86	15.6	1693	14	HEPATITIS E VIRUS COM	6.27e-01
8	86	15.6	1693	14	UNNAMED PROTEIN PRODUC	6.27e-01
9	86	15.6	1693	14	METHYL TRANSFERASE.	6.27e-01
10	84	15.2	622	2	CRYIIC DELTA-ENDOTOXIN	1.21e+00
11	84	15.2	921	5	C34B7.2 PROTEIN.	1.21e+00
12	83	15.1	180	10	PUTATIVE ETHYLENE RECE	1.68e+00
13	83	15.1	281	9	ORF 1.	1.68e+00
14	83	15.1	1577	2	GLUCOSYLTRANSFERASE.	1.68e+00
15	83	15.1	1693	14	NONSTRUCTURAL POLYPROT	1.68e+00
16	81	14.7	196	14	PUTATIVE COAT PROTEIN.	3.21e+00
17	81	14.7	216	2	FLAGELLAR L-RING PROTE	3.21e+00
18	81	14.7	633	2	INSECTICIDAL CRYSTAL P	3.21e+00
19	81	14.7	642	14	PUTATIVE READTHROUGH P	3.21e+00
20	81	14.7	3011	14	GENOME POLYPROTEIN.	3.21e+00

21	80	14.5	275	14	Q68467	POLYPROTEIN (FRAGMENT)	4.41e+00
22	80	14.5	275	14	Q68468	POLYPROTEIN (FRAGMENT)	4.41e+00
23	80	14.5	275	14	Q68469	POLYPROTEIN (FRAGMENT)	4.41e+00
24	80	14.5	357	5	Q01993	VC27A7L.1 PROTEIN	4.41e+00
25	80	14.5	802	3	Q07034	RNA BINDING PROTEIN.	4.41e+00
26	80	14.5	2436	14	Q81756	POLYPROTEIN (FRAGMENT)	4.41e+00
27	79	14.3	370	3	Q04083	D9461.24P.	6.06e+00
28	79	14.3	542	5	Q17456	SIMILAR TO GLUTAMATE D	6.06e+00
29	79	14.3	860	1	Q59003	860AA LONG HYPOTHETICA	6.06e+00
30	78	14.2	70	2	P75688	FROM BASES 311709 TO 3	8.28e+00
31	78	14.2	613	10	Q38846	ETHYLENE RESPONSE SENS	8.28e+00
32	78	14.2	3011	14	Q36609	POLYPROTEIN.	8.28e+00
33	78	14.2	3011	14	Q36608	POLYPROTEIN.	8.28e+00
34	78	14.2	3011	14	Q36610	POLYPROTEIN.	8.28e+00
35	78	14.2	3011	14	Q36579	POLYPROTEIN.	8.28e+00
36	77	14.0	418	2	Q56631	LECITHINASE.	1.13e+01
37	77	14.0	470	2	Q87325	LECITHINASE.	1.13e+01
38	77	14.0	510	10	Q65322	PUTATIVE MONOSACCHARID	1.13e+01
39	77	14.0	967	2	Q54123	PEPB.	1.13e+01
40	77	14.0	1361	2	Q86617	HYPOTHETICAL 145.0 KD	1.13e+01
41	77	14.0	1464	11	Q08948	N-METHYL-D-ASPARTATE R	1.13e+01
42	77	14.0	1464	11	Q63728	N-METHYL-D-ASPARTATE R	1.13e+01
43	77	14.0	1464	4	Q12879	N-METHYL-D-ASPARTATE R	1.13e+01
44	77	14.0	2219	5	Q23388	ZK1067.2 PROTEIN.	1.13e+01
45	76	13.8	821	2	Q51735	OUTER MEMBRANE PROTEIN	1.54e+01

ALIGNMENTS

RESULT 1
ID Q81876 PRELIMINARY; PRT: 1693 AA.
AC Q81876;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE COMPLETE GENOME SEQUENCE.
OS HEPATITIS E VIRUS (HEV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALCIVIRIDAE;
RN CALICIVIRUS.
RC SEQUENCE FROM N.A.
RA UCHIDA T.;
RL SUBMITTED (MAY-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: D11093; D1002342;
SQ SEQUENCE 1693 AA; 184840 MW; 8952DD38 CRC32;

Query Match 18.1%; Score 100; DB 14; Length 1693;
Best Local Similarity 26.9%; Pred. No. 4.78e-03;
Matches 18; Conservative 22; Mismatches 24; Indels 3; Gaps 3;

Db	572	FRTSFVDGAVLEANGPERYNLSFDASQSTMAAGPFLSYAASAGLEVRYVAAGLDHRAV	631
Qy	8	FGQGYVTPFLSESNVRYKISAGSCPLSTAGPSYVFKQDNPNVGSOT-F-SAGLHLR-V	64
Db	632	FAPGVSP	638
Qy	65	FDPSTGA	71

RESULT 2			
ID	O41088	PRELIMINARY; PRT: 114 AA.	
AC	O41088;		
DT	01-JAN-1998 (TREMBLREL. 05, CREATED)		
DT	01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)		
DE	A606L PROTEIN.		
GN	A606L.		
OS	PARAMECIDIUM BURSARIA CHLORELLA VIRUS 1 (PBCV-1).		
OC	VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PHYCODNAVIRIDAE; PHYCODNAVIRUS.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE; 98022962.		

```
CC  -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC -----
CC  EMBL: Z36098; G536526; -
CC  PIR: S46105; S46105.
CC  PROSITE: PS00129; GLYCOSYL_HYDROL_F31_1; 1.
CC  PROSITE: PS00707; GLYCOSYL_HYDROL_F31_2; FALSE_NEG.
CC  PFAM: PF01055; Glycosyl_hydrl5; 1.
CC  KW  HYPOTHETICAL PROTEIN: HYDROLASE; GLYCOSIDASE.
CC  FT  ACT_SITE 537
CC  SQ  SEQUENCE 954 AA; 110265 MW; 9710BA62 CRC32;
CC -----
Query Match 14.5%; Score 80; DB 1; Length 954;
Best Local Similarity 21.5%; Pred. No. 1.59e+00;
Matches 14; Conservative 20; Mismatches 29; Indels 2; Gaps 2;
Db 136 KFOEEANRTSIPQHFLLKQQTVNSFWSKISSFLSLNSTADTFHLRNGDYSVEIFAEPF 195
QY 3 KFAAEFGQGVV-QTPFLSESNVR-YKISAGSCPLSTAGPSYVKFQDNPVGSQTSAGL 60
Db 196 OLKVV 200
QY 61 HLRVF 65
Search completed: Wed Sep 1 16:12:47 1999
Job time : 15 secs.
```

```
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE    NUCLEAR POLYADENYLATED RNA-BINDING PROTEIN GN33.
GN    NAB3 OR YPLI90C.
OS    SACCHAROMYCETES CEREVISIAE (BAKER'S YEAST).
OC    EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCUMYCETES; SACCHAROMYCETALES;
SC    SACCHAROMYCETACEAE; SACCHAROMYCETES.
[.] 
RN    SEQUENCE FROM N.A.
RP    WILSON S.M., OBERDORF A.M., DATAR K.V., SWEDLOW J.R., PADDY M.R.,
RA    SWANSON M.S.;
RL    SUBMITTED (JAN-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
[+] 
RN    SEQUENCE FROM N.A.
RP    RIEGER M.T., MOELLER-AUER S., SCHAEFER M.;
RL    SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC    -1- SUBCELLULAR LOCATION: NUCLEAR.
CC    -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNP).
-----
CC    This swiss-prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outpost at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to licensel@isb-sib.ch).
-----
DR    EMBL; U05314; G476220; -.
DR    EMBL; Z73546; E246917; -.
DR    PIR; S48529; S48529.
DR    SGD; L0001228; NAB3.
DR    PFAM; PF00076; rtm: 1.
KW    RNA-BINDING; NUCLEAR PROTEIN.
FT    DOMAIN      68       71     POLY-GLU.
FT    DOMAIN      87       93     POLY-LYS.
FT    DOMAIN     101      106     POLY-ASP.
FT    DOMAIN     108      115     POLY-LYS.
FT    DOMAIN     116      127     POLY-GLU.
FT    DOMAIN     128      137     POLY-ASP.
FT    DOMAIN     137      147     POLY-LYS.
FT    DOMAIN     151      160     POLY-LYS.
FT    DOMAIN     160      168     POLY-PRO.
FT    DOMAIN     168      178     POLY-LYS.
FT    DOMAIN     178      188     POLY-GLN.
FT    DOMAIN     188      203     POLY-PRO.
FT    DOMAIN     203      218     POLY-GLN.
FT    DOMAIN     218      228     POLY-ASP.
FT    DOMAIN     228      238     POLY-PRO.
FT    DOMAIN     238      248     POLY-GLN.
SQ    SEQUENCE 802 AA; 90438 MW; FB180EDB CRC32;
```

Query Match 14.5%; Score 80; DB 1; Length 802;
Best Local Similarity 32.6%; Pred. No. 1.59e+00;
Matches 14; Conservative 10; Mismatches 16; Indels 3; Gaps 3;

```
Dd    622 QGYGSPPIPMNSYG-RYTSTISPPPPPQQQIQGVGRVCGAP 663  
          ||| |.:|||.:.||||.|:||..||.:|||.:.  
Qy    10 QGY-VQTFF-LSSSNRYKIISIAGSCPPLSTAGPSVKFDNP 50
```

RESULT 15

ID	YB79_YEAST	STANDARD;	PRT:
AC	F38138;		954 AA.
DT	01-OCT-1994 (REL. 30, CREATED)		
DT	01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)		
DT	01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)		
DE	POTATIVE FAMILY 31 GLUCOSIDASE IN PCS60-ABDI INTERGENIC REGION (EC 3.2.1.-).		
GN	YBR229C OR YBR1526.		
OS	SACCHAROMYCETES CEREVISIAE (BAKER'S YEAST).		
OC	EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCUMYCETES; SACCHAROMYCETALES;		
OC	SACCHAROMYCETACEAE; SACCHAROMYCETES.		
RN	[.]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-S288C;		
RA	DUBOIS B., EL BAKKOURY M., GLANSDOUFF N., MESSENGUY F., PIERARD A.,		
RI	SCHREIBERS B., VIERENDIELS F.;		
RU	SUBMITTED (AUG-1994) TO EMBL / GENBANK / DDBJ DATA BANKS.		

```
FT CONFLICT 45 45 H -> R (IN REF. 2).
SQ SEQUENCE 527 AA; 61239 MW; 8DF019E3 CRC32;

Query Match
Best Local Similarity 44.4%; Pred. No. 7.96e-01;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 242 LRYVGLMGSLCPATGFSYKWEIDP 268
QY 24 VRYKISAGSCLPSTAGPSYKVFQDNP 50

RESULT 11
ID TAX_MOUSE STANDARD; PRT; 527 AA.
AC P42682;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE TXK (EC 2.7.1.112) (PTK-RL-18) (RESTING
LN LYMPHOCYTE KINASE).
GN TXK OR RLK.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=THYMUS;
RX MEDLINE; 96059536.
RA HAIRE R.N., LITMAN G.W.;
RT "The murine form of TXK, a novel TEC kinase expressed in thymus maps
RT to chromosome 5.",
RL MAMM. GENOME 6:476-480(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=THYMUS;
RA SOMMERS C.L., HUANG K., GRINBERG A., CHARLICK D.A., KOZAK C.A.,
RA LOVE P.E.;
RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=LIVER;
RA HIGASHITSUJI H., NONOGUCHI K., ARII S., FURUTANI M., KANEKO Y.,
RA NAKAYAMA H., FUJITA J.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=THYMUS;
RX MEDLINE; 95130578.
RA HU O., DAVIDSON D., SCHWARTZBERG P.L., MACCHIARINI F.,
RA LENARDO M.J., BLUESTONE J.A., MATIS L.A.;
RT "Identification of RLK, a novel protein tyrosine kinase with
RT predominant expression in the T cell lineage.",
RL J. BIOL. CHEM. 270:1928-1934(1995).
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN EARLY THYMOCYTES, T CELLS AND
CC MAST CELLS.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN. BELONGS TO THE BTK SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U16145; G562125; -;
DR EMBL; U19607; G643065; -;
```

```
DR EMBL; D43963; G604884; -;
DR EMBL; L35268; G623443; -;
DR MGD; MGI:102960; TXK.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00001; SH2; 1.
DR PROSITE; PS00002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSP; Q06187; IAWW.
KW TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;
KW SH3 DOMAIN; PHOSPHORYLATION.
FT DOMAIN 14 20 POLY-CYS.
FT DOMAIN 82 142 SH3.
FT DOMAIN 150 246 SH2.
FT DOMAIN 271 527 PROTEIN KINASE.
FT NP_BIND 277 285 ATP (BY SIMILARITY).
FT BINDING 299 299 ATP (BY SIMILARITY).
FT ACT_SITE 390 390 BY SIMILARITY.
FT MOD_RES 420 420 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CONFLICT 3 4 LS -> SF (IN REF. 3).
FT CONFLICT 6 6 Y -> D (IN REF. 3).
FT CONFLICT 272 272 A -> T (IN REF. 3).
FT CONFLICT 497 497 R -> S (IN REF. 3).
SQ SEQUENCE 527 AA; 61108 MW; 5B39DA78 CRC32;

Query Match
Best Local Similarity 44.4%; Pred. No. 1.13e+00;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 242 LRYPIGLGSLCPATGFSYKWEIDP 268
QY 24 VRYKISAGSCLPSTAGPSYKVFQDNP 50

RESULT 12
ID YCB2_PSEDE STANDARD; PRT; 171 AA.
AC P29944;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 19.0 KD PROTEIN IN COBS 5'REGION (ORF2).
OS PSEUDOMONAS DENITRIFICANS.
OC BACTERIA; PROTEOBACTERIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92011364.
RA CAMERON B., GUILHOT C., BLANCHE F., CAUCHOIS L., ROUYEZ M.-C.,
RA RIGAUD S., LEVY-SCHIL S., CROUZET J.;
RT "Genetic and sequence analyses of a Pseudomonas denitrificans DNA
RT fragment containing two cob genes.",
RL J. BACTERIOL. 173:6058-6065(1991).
CC -!- SIMILARITY: CONTAINS A DNAJ-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M62869; G151167; -;
DR PIR; B38162; B38162.
DR PROSITE; PS00636; DNAJ_1; FALSE_NEG.
DR PROSITE; PS00076; DNAJ_2; 1.
DR PFAM; PF00226; DnaJ; 1.
DR HSP; P08622; 1XBL.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 171 AA; 18973 MW; 5DDD21D5 CRC32;
```



```

CC EMBL; M80581; G329998;
KW POLYPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; HELICASE;
FT NP-BINDING. 975 982 ATP (POTENTIAL).
SQ SEQUENCE 1693 AA; 185149 MW; FBCA2483 CRC32;

Query Match 15.6%; Score 86; DB 1; Length 1693;
Best Local Similarity 25.4%; Pred. No. 1.92e-01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 FRTSVDGAVLETNGPERHNLSFASQSTMAAGPFSLTYSAAAGLEGVRYVYAAGLDHRV 631
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 8 FGQYVQTPFLSESNVRYKISAGSCLSTAGPSYVKFQDNPVGSQT-F-SAGLHLR-V 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 632 FAPGVSP 638
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 65 FDPSTGA 71
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
ID Y4II_RHISN STANDARD; PRT; 703 AA.
AC P55492;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 76.2 KD PROTEIN Y4II.
GN Y4II.
OS RHIZOBIUM SP. (STRAIN NGR234).
OG PLASMID SYN PNGR234A.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
OC RHIZOBIACEAE; RHIZOBIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97305956.
RA FREIBERG C.A., FELLAY R., BAIROCH A., BROUGHTON W.J., ROSENTHAL A.,
RA PERRET X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL NATURE 387:394-401(1997).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: NONE OBVIOUS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000078; G2182446;
KW HYPOTHETICAL PROTEIN; PLASMID; TRANSMEMBRANE.
FT TRANSMEM 23 43 POTENTIAL.
FT TRANSMEM 69 89 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 250 270 POTENTIAL.
FT TRANSMEM 357 377 POTENTIAL.
FT TRANSMEM 432 452 POTENTIAL.
FT TRANSMEM 644 664 POTENTIAL.
SQ SEQUENCE 703 AA; 76183 MW; A2BA53CE CRC32;

Query Match 15.4%; Score 85; DB 1; Length 703;
Best Local Similarity 27.0%; Pred. No. 2.75e-01;
Matches 17; Conservative 17; Mismatches 28; Indels 1; Gaps 1;

Db 224 SPADTDSGTISIATSLFGAGASTRPEWLAFAANALASGAPFGLSPRVATATSOYA 283
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 15 TPFL-SESNVRYKISAGSCLSTAGPSYVKFQDNPVGSQTF-SAGLHLRVPDPTGALV 73
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 284 KAK 286
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 74 DSK 76
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 7
ID CPAC_ECOLI STANDARD; PRT; 869 AA.
AC P25733;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE CFA/I FIMBRIAL SUBUNIT C PRECURSOR (COLONISATION FACTOR ANTIGEN I
DE SUBUNIT C).
GN CPAC.
OS ESCHERICHIA COLI.
OG PLASMID NTP513.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-ENTEROTOXIGENIC;
RX MEDLINE; 89330163.
RA HAMERS A.M., PEL H.J., WILLSHAW G.A., KUSTERS J.G.,
RA VAN DER ZEIJST B.A.M., GAASTRA W.;
RT "The nucleotide sequence of the first two genes of the CFA/I fimbrial
RT operon of human enterotoxigenic Escherichia coli.";
RL MICROB. PATHOG. 6:297-309(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92329981.
RA JORDI B.J.A.M., WILLSHAW G.A., VAN DER ZEIJST B.A.M., GAASTRA W.;
RT "The complete nucleotide sequence of region 1 of the CFA/I fimbrial
RT operon of human enterotoxigenic Escherichia coli.";
RL DNA SEQ. 2:257-263(1992).
CC -1- FUNCTION: MAY SERVE AS ANCHOR FOR THE FIMBRIAE IN THE OUTER
CC MEMBRANE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M55661; G145510;
KW ANTIGEN; SIGNAL; FIMBRIA; OUTER MEMBRANE; PLASMID.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 869 CFA/I FIMBRIAL SUBUNIT C.
SQ SEQUENCE 869 AA; 97830 MW; 7AF76347 CRC32;

Query Match 15.4%; Score 85; DB 1; Length 869;
Best Local Similarity 32.4%; Pred. No. 2.75e-01;
Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;

Db 158 AFQSQTINLSDSKYKRLSISGNSALGITDTSY 191
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 11 GYVQTPFLSESNVRYK-ISIAGSCLSTAGPSY 43
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
ID YX45_MYCTU STANDARD; PRT; 611 AA.
AC Q10830;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 67.2 KD PROTEIN CY274.45C.
GN MYCY274.45C OR MYCY338.01C.
OS MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FRMICTUTES; ACTINOBACTERIA; ACTINOBACTERIACEAE;
OC ACTINOMYCETALES; CORYNEBACTERIACEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-H37RV;
RA CONNOR R., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: BELONGS TO THE N-ACYL-D-AMINO-ACID DEACYLASE FAMILY.

```

```
CC ----- TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10330; G221705; -
KW POLYPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; HELICASE;
KW ATP-BINDING.
FT NP_BIND 975 982 ATP (POTENTIAL).
SQ SEQUENCE 1693 AA; 185215 MW; FFCB786D CRC32;
Query Match 18.1%; Score 100; DB 1; Length 1693;
Best Local Similarity 26.9%; Pred. No. 9.43e-04; Indels 3; Gaps 3;
Matches 18; Conservative 22; Mismatches 24;
Db 572 FRTSFVDGAVLEANGPERYNLSFDASQSTMAAGPFSLTAAAGLEVRVYVAAGLDHRV 631
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 8 FGQGYVQTPFLSESNVRYKISAGSCPLSTAGPSYKVFQDNVPGSQT-F-SAGLHLR-V 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 632 FAGGVSP 638
| : : :
QY 65 FDPSTGA 71
| : : :
RESULT 3
ID EMB8_PICGL STANDARD; PRT; 457 AA.
AC Q40863;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE LATE EMBRYOGENESIS ABUNDANT PROTEIN EMB8.
GN EMB8.
OS PICEA GLAUCA (WHITE SPRUCE).
OC EUKARYOTA; VIRIDIPALANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; CONIFEROPSIDA; CONIFERALES; PINACEAE;
OC PICEA.
RN [1]
RP SEQUENCE FROM N.A.
RA DONG J.Z., DUNSTAN D.I.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: BELONGS TO THE UPF0017 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L47118; G1350545; -
DR PROSITE; PS01133; UPF0017; 1.
SQ SEQUENCE 457 AA; 51019 MW; AE7CB4CD CRC32;
Query Match 15.6%; Score 86; DB 1; Length 457;
Best Local Similarity 34.0%; Pred. No. 1.92e-01; Indels 2; Gaps 2;
Matches 17; Conservative 11; Mismatches 20;
Db 232 LGANILVRYLGEVAGNCPLSGAVSLNPF-NLVIADDFHKGGLGFNNVYD 280
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 18 LSESNVRYKISAGSCPLSTAGPSYKVFQDNVPGSQTFSAGLHLR-VFD 66
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 4
ID POLN_HEVBU STANDARD; PRT; 1693 AA.
AC P29324;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
```

```
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE NON-STRUCTURAL POLYPROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE
DE (EC 2.7.7.48); HELICASE].
OS HEPATITIS E VIRUS (STRAIN BURMA) (HEV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALICIVIRIDAE;
OC CALICIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92024067.
RA TAM A.W., SMITH M.M., GUERRA M.E., HUANG C.-C., BRADLEY D.W.,
RA FRY K.E., REYES G.R.;
RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT full-length viral genome."
RL VIROLOGY 185:120-131(1991).
CC -!- HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF ENTERICALLY
CC TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M73218; G330024; -
DR PIR: A40778; MNWHE.
KW POLYPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; HELICASE;
KW ATP-BINDING.
FT NP_BIND 975 982 ATP (POTENTIAL).
SQ SEQUENCE 1693 AA; 185191 MW; C560BE14 CRC32;
Query Match 15.6%; Score 86; DB 1; Length 1693;
Best Local Similarity 25.4%; Pred. No. 1.92e-01; Indels 3; Gaps 3;
Matches 17; Conservative 22; Mismatches 25;
Db 572 FRTSFVDGAVLENGPERHNLSDASQSTMAAGPFSLTAAAGLEVRVYVAAGLDHRV 631
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 8 FGQGYVQTPFLSESNVRYKISAGSCPLSTAGPSYKVFQDNVPGSQT-F-SAGLHLR-V 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 632 FAGGVSP 638
| : : :
QY 65 FDPSTGA 71
| : : :
RESULT 5
ID POLN_HEVPA STANDARD; PRT; 1693 AA.
AC P33424;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE NON-STRUCTURAL POLYPROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE
DE (EC 2.7.7.48); HELICASE].
OS HEPATITIS E VIRUS (STRAIN PAKISTAN) (HEV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALICIVIRIDAE;
OC CALICIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92115700.
RA TSAREV S.A., EMERSON S.U., REYES G.R., TSAREVA T.S., LECTERS L.J.,
RA MALIK I.A., IQBAL M., PURCELL R.H.;
RT "Characterization of a prototype strain of hepatitis E virus."
RL PROC. NATL. ACAD. SCI. U.S.A. 89:559-563(1992).
CC -!- HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF ENTERICALLY
CC TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

(TM)

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	300	54.4	65	1	Y15A_BP74	HYPOTHETICAL 7.3 KD PR	6.06e-47
2	100	18.1	1693	1	POLN_HEVW	NON-STRUCTURAL POLYPRO	9.43e-04
3	86	15.6	457	1	EMB8_PICGL	LATE EMBRYOGENESIS ABU	1.92e-01
4	86	15.6	1693	1	POLN_HEVPA	NON-STRUCTURAL POLYPRO	1.92e-01
5	86	15.6	1693	1	POLN_HEVPA	NON-STRUCTURAL POLYPRO	1.92e-01
6	85	15.4	703	1	Y41I_RHLSN	HYPOTHETICAL 76.2 KD P	2.75e-01
7	85	15.4	869	1	CFAC_ECOLI	CEA/I FIMBRIAL SUBUNIT	2.75e-01
8	83	15.1	611	1	YX45_MYCTU	HYPOTHETICAL 67.2 KD P	5.61e-01
9	82	14.9	449	1	YM60_YEAST	HYPOTHETICAL 51.4 KD P	7.96e-01
10	82	14.9	527	1	TKX_HUMAN	TYROSINE-PROTEIN KINAS	7.96e-01
11	81	14.7	527	1	TKX_HUMAN	TYROSINE-PROTEIN KINAS	1.13e+00
12	80	14.5	571	1	YCB2_PSEDE	HYPOTHETICAL 19.0 KD P	1.59e+00
13	80	14.5	172	1	YK42_TRVBB	RECEPTOR-TYPE ADENYLAT	1.59e+00
14	80	14.5	802	1	NAB3_YEAST	NUCLEAR POLYADENYLATED	1.59e+00
15	80	14.5	954	1	YB7_YEAST	PUTATIVE FAMILY 31 GLU	1.59e+00
16	80	14.5	1097	1	IMB3_HUMAN	IMPORTIN BETA-3 SUBUNIT	1.59e+00
17	80	14.5	1659	1	VIT1_ONCMY	VITELLOGENIN PRECURSOR	1.59e+00
18	80	14.5	3011	1	POLG_HCVI	GENOME POLYPROTEIN [CO	1.59e+00
19	79	14.3	472	1	VTDB_MOUSE	VITAMIN D-BINDING PROT	2.24e+00
20	79	14.3	474	1	VTDB_HUMAN	VITAMIN D-BINDING PROT	2.24e+00
21	79	14.3	633	1	CR72_BACTK	70 KD CRYSTAL PROTEIN	2.24e+00
22	79	14.3	935	1	YJ31_YEAST	HYPOTHETICAL 108.4 KD	2.24e+00
23	78	14.2	355	1	YQHT_BACSU	PUTATIVE PEPTIDASE IN	3.14e+00


```

Best Local Similarity 44.4%; Pred.No.2.96e+00;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 242 LRYPIGLLGSCLPATSGFSYKWEIDP 268
    :: ||| :|| :|| |:: :|
Qy 24 VRYKISAGSCPSTAGPSVYKFQDNP 50
    :: ||| :|| :|| |:: :|

RESULT 14
ENTRY Txk - mouse #type complete
TITLE Txk - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
      12-Feb-1999
ACCESSIONS I49133
REFERENCE I49133
          Haire, R.N.; Litman, G.W.
#journal Mamm. Genome (1995) 6:476-480
#title The murine form of TXK, a novel TEC kinase expressed in
        thymus maps to chromosome 5.
#cross-references MUID:96059536
#accession I49133
#status preliminary; translated from GB/EMBL/DDBJ
##molecule_type mRNA
##residues 1-527 ##label RES
##cross-references EMBL:U16145; NID:g562124; PID:g562125
CLASSIFICATION #superfamily protein-tyrosine kinase tec; pleckstrin repeat
                homology; protein kinase homology; SH2 homology; SH3
                homology
KEYWORDS ATP
FEATURE
89-137 #domain SH3 homology #label SH3\
150-246 #domain SH2 homology #label SH2\
269-527 #domain protein kinase homology #label KIN\
277-285 #region protein kinase ATP-binding motif
SUMMARY #length 527 #molecular-weight 61108 #checksum 785

Query Match 14.7%; Score 81; DB 2; Length 527;
Best Local Similarity 44.4%; Pred.No.2.96e+00;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 242 LRYPIGLLGSCLPATSGFSYKWEIDP 268
    :: ||| :|| :|| |:: :|
Qy 24 VRYKISAGSCPSTAGPSVYKFQDNP 50
    :: ||| :|| :|| |:: :|

RESULT 15
ENTRY S40770 #type complete
TITLE polyprotein precursor - hepatitis C virus
ORGANISM #formal_name hepatitis C virus
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
      18-Sep-1998
ACCESSIONS S40770; PC1285
REFERENCE S40770
          Okamoto, H.
#authors submitted to the EMBL Data Library, March 1992
#submission S40770
#accession S40770
#status preliminary
##molecule_type genomic RNA
##residues 1-3011 ##label OKA
##cross-references EMBL:D10749; NID:g221586; PID:d1002057; PID:g221587
PC1284
#authors Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka,
          T.; Yoshizawa, H.; Tsuda, F.; Miyakawa, Y.; Mayumi, M.
#journal J. Exp. Med. (1990) 60:167-177
#title The 5'-terminal sequence of the hepatitis C virus genome.
#accession PC1285
##molecule_type genomic RNA
##residues 1-513 ##label OK2
##experimental_source isolate HC-J1
##classification #superfamily hepatitis C virus genome polyprotein; DEAD/H box
                helicase homology
                polyprotein
KEYWORDS

```

```

RESULT 12
ENTRY flagellar L-ring protein flgH - Aquifex aeolicus
TITLE flagellar L-ring protein flgH - Aquifex aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change 08-May-1998

ACCESSIONS G70447
REFERENCE A70300
#authors Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; Aulay, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
#cross-references MUID:98196666
#accession G70447
#status preliminary; nucleic acid sequence not shown; translation not shown

##molecule_type DNA
##residues 1-216 #label AQF
##cross-references GB:AE000752; NID:g2984021; PID:g2984027; GB:AE000657
##experimental_source strain VF5

GENETICS
#gene flgH
#length 216 #molecular-weight 23767 #checksum 9605

SUMMARY
Query Match 14.7% Score 81; DB 2; Length 216;
Best Local Similarity 27.7% Pred. No. 2.96e+00;
Matches 18; Conservative 25; Mismatches 16; Indels 6; Gaps 5

Db 93 SSFGVHPAILK-NLGAGGKSSPASKGGS--KQQSQVLTTTL-AGRVKVF-PNGTMLV 147
::: : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 15 TPFLLS-ESNSVRYKISITAGSCPLSTAGPSYKFDQNPVGSQTFSAGLHLRFVDPSTGALV 73
::: : : : : : : : : : : : : : : : : : : : : : : : :

Db 148 EAKKY 152
::: : : : : : : : : : : : : : : : : : : : : : : : :
Qy 74 DSKSY 78

RESULT 13
ENTRY protein-tyrosine kinase (EC 2.7.1.112) rlk - mouse
TITLE resting lymphocyte kinase
ALTERNATE_NAMES #formal_name Mus sp. #common_name mouse
ORGANISM 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 17-Mar-1999
DATE #cross-references MUID:95130578

ACCESSIONS A55631
REFERENCE A55631
#authors Hu, Q.; Davidson, D.; Schwartzberg, P.L.; Macchiarini, F.; Lenardo, M.J.; Bluestone, J.A.; Matis, L.A.
#journal J. Biol. Chem. (1995) 270:1928-1934
#title Identification of rlk, a novel protein tyrosine kinase with predominant expression in the T cell lineage.
#cross-references MUID:95130578
#accession A55631
#status preliminary
##molecule_type mRNA
##residues 1-527 #label HUA
##cross-references GB:L35268; NID:g623442; PID:g623443
CLASSIFICATION #superfamily protein-tyrosine kinase tec; pleckstrin repeat homology; protein kinase homology; SH2 homology; SH3 homology
phosphotransferase

KEYWORDS
FEATURE
89-137 #domain SH3 homology #label SH3\
150-246 #domain SH2 homology #label SH2\
269-527 #domain protein kinase homology #label KIN\
277-285 #region protein kinase ATP-binding motif
SUMMARY #length 527 #molecular-weight 61108 #checksum 785

Query Match 14.47% Score 81; DB 2; Length 527;

```

```
ALTERNATE_NAMES      plasmid
delta-endotoxin
#formal_name Bacillus thuringiensis
#sequence_revision 10-Nov-1995 #text_change
DATE 22-Nov-1993
ACCESSIONS          S17400
REFERENCE            S17402
#authors            Wu, D.; Cao, X.L.; Bai, Y.Y.; Aronson, A.I.
#journal            FEMS Microbiol. Lett. (1991) 81:31-36
#title              Sequence of an operon containing a novel delta-endotoxin gene
                    from Bacillus thuringiensis.
#accession          S17402
##status            preliminary
##molecule_type    DNA
##residues          1-622 #label WUD
##cross-references  EMBL:X57252; NID:g40283; PID:g40286
GENETICS
#genome             plasmid
#keywords            delta-endotoxin
SUMMARY
#length 622 #molecular-weight 69729 #checksum 8205
Query Match          15.2%; Score 84; DB 2; Length 622;
Best Local Similarity 18.2%; Pred. No. 1.13e+00;
Matches 12; Conservative 26; Mismatches 26; Indels 2; Gaps 2;
Db 501 FISEKYNQDLSRFELSNPTARYTLRGNNSYNLRLVSSIGSSIRVTINGRVYTNV 560
QY 12 YVOTPELSENSVRYKISAGS-CPLSTAGPSYVKF-QDNPVGSQTFSGAGLHLRVDPST 69
Db 561 NTTTTN 566
QY 70 GALVDS 75

RESULT 7
ENTRY S38913 #type complete
TITLE hypothetical protein 1 - phage phi-C31
ORGANISM #formal_name phage phi-C31
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
ACCESSIONS          S38913
REFERENCE            S38912
#authors            Hartley, N.M.; Murphy, G.O.; Bruton, C.J.; Chater, K.F.
#submission          submitted to the EMBL Data Library, November 1993
#accession           S38913
##status            preliminary
##molecule_type    DNA
##residues          1-281 #label HAR
##cross-references  EMBL:X76288; NID:g432610; PID:g579071
GENETICS
#start_codon        GTG
SUMMARY
#length 281 #molecular-weight 31680 #checksum 2790
Query Match          15.1%; Score 83; DB 2; Length 281;
Best Local Similarity 24.1%; Pred. No. 1.56e+00;
Matches 14; Conservative 17; Mismatches 24; Indels 3; Gaps 3;
Db 194 AYMNADFIIDPGNRPMPFEFGAAVLHYVDETWA-FKPVTGPDVFAQFLHLRQTFD 250
QY 11 GYVQTPFLSENSVRYKIS-IAGSCPLSTAGPSYVKFQDNPVGSQTFSGAGLHLR-VFD 66

RESULT 8
ENTRY D70928 #type complete
TITLE hypothetical protein Rv2913c - Mycobacterium tuberculosis
                    (strain H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
ACCESSIONS          D70928
REFERENCE            A70500
#authors            Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
                    C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
                    III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
                    Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
                    Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, K.;
                    Horsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
                    Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
                    Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
                    Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
                    Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal            Nature (1998) 393:537-544
#title              Deciphering the biology of Mycobacterium tuberculosis from
                    the complete genome sequence.
#cross-references  MUID:98295987
#accession          D70928
##status            preliminary; nucleic acid sequence not shown;
                    translation not shown
##molecule_type    DNA
##residues          1-611 #label COL
##cross-references  GB:274024; GB:AL123456; NID:g3250700; PID:el301028;
                    PID:g3250707
##experimental_source strain H37Rv
GENETICS
#gene              Rv2913c
#summary            #length 611 #molecular-weight 67204 #checksum 9971
Query Match          15.1%; Score 83; DB 2; Length 611;
Best Local Similarity 43.5%; Pred. No. 1.56e+00;
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
Db 319 VRFOHLPVPFELYDGDIDLVPFE 341
QY 44 VKFQDNPVGSQTFSGAGLHLRVFD 66

RESULT 9
ENTRY I38375 #type fragment
TITLE tyrosine kinase - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 29-May-1998 #sequence_revision 29-May-1998 #text_change
ACCESSIONS          I38375
REFERENCE            I38372
#authors            Haire, R.N.; Ohta, Y.; Lewis, J.E.; Fu, S.M.; Kroisel, P.;
                    Litman, G.W.
#journal            Hum. Mol. Genet. (1994) 3:897-901
#title              TKX, a novel human tyrosine kinase expressed in T cells
                    shares sequence identity with Tec family kinases and maps
                    to 4p12.
#cross-references  MUID:95038742
#accession          I38375
##status            preliminary; translated from GB/EMBL/DBJ
##molecule_type    DNA
##residues          1-81 #label RES
##cross-references  EMBL:U07794; NID:g508219; PID:g508224
GENETICS
#gene              GDB:TXK
#map_position        4p12-4p12
#introns             25/1
CLASSIFICATION
#superfamily         protein-tyrosine kinase tec; pleckstrin repeat
                    homology; protein kinase homology; SH2 homology; SH3
                    homology
FEATURE
32-81 #domain protein kinase homology (fragment) #label KIN
SUMMARY
#length 81 #checksum 7751
Query Match          14.9%; Score 82; DB 2; Length 81;
Best Local Similarity 44.4%; Pred. No. 2.15e+00;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
Db 5 LRPVGLMGSLPATAGSYKWEIDP 31
QY 24 VRYKISAGSCPLSTAGPSYVKFQDNP 50
```

```
#cross-references MUID:99000809
#accession G71525
##status preliminary
##molecule_type DNA
##residues 1-566 ##label ARN
##cross-references GB:AE001308; GB:AE001273; NID:g3328766; PID:g3328771
##experimental_source serotype D, strain UW-3/Cx
GENETICS
#gene CT350
SUMMARY
#length 566 #molecular-weight 63507 #checksum 4960
Query Match 16.2%; Score 89; DB 2; Length 566;
Best Local Similarity 26.2%; Pred. No. 2.14e-01;
Matches 17; Conservative 25; Mismatches 18; Indels 5; Gaps 5;
Db 63 IAEYLOQSFLEDTYIR-KSAIGA-GLSGSSEA-LELLSEATQDLYEQL-L-ILNA 117
QY 8 FGQGYVQTPFLSESNVRYKISAGSCLPLTAGPSYVKFQDNPVGSQTFSGAGLHLRVDP 67
Db 118 ATSQL 122
QY 68 STGAL 72
RESULT 3
ENTRY MNWHE #type complete
CONTAINS genome polyprotein - hepatitis E virus (strain Burma)
ORGANISM RNA-directed RNA polymerase (BC 2.7.7.48)
#formal_name hepatitis E virus
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
ACCESSIONS A40778; A48547
REFERENCE A40778
#authors Tam, A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes, G.R.
#journal Virology (1991) 185:120-131
#title Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome.
#cross-references MUID:92024067
#accession A40778
##residues 1-1693 ##label TAM
##molecule_type genomic RNA
##cross-references GB:M73218; NID:g330023; PID:g330024
REFERENCE A48547
#authors Fry, K.E.; Tam, A.W.; Smith, M.M.; Kim, J.P.; Luk, K.C.; Young, L.M.; Piatlak, M.; Feldman, R.A.; Yun, K.Y.; Purdy, M.A.; McCaustland, K.A.; Bradley, D.W.; Reyes, G.R.
#journal Virus Genes (1992) 6:173-185
#title Hepatitis E virus (HEV): strain variation in the nonstructural gene region encoding consensus motifs for an RNA-dependent RNA polymerase and an ATP/GTP binding site.
#cross-references MUID:92271462
#accession A48547
##molecule_type genomic RNA
##residues 967-1693 ##label FRY
##cross-references GB:M32400; NID:g330021; PID:g330022
##note sequence extracted from NCBI backbone (NCBIN:104572, NCBIP:104573)
CLASSIFICATION #superfamily hepatitis E virus nonstructural protein
KEYWORDS ATP; nonstructural protein; nucleotidyltransferase
SUMMARY #length 1693 #molecular-weight 185191 #checksum 6520
Query Match 15.6%; Score 86; DB 1; Length 1693;
Best Local Similarity 25.4%; Pred. No. 5.84e-01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;
Db 572 FRTSFVDGAVLENGPERNLSPDASQSTWAGPFFSLTYAASAGLEVRYVAGLDRHV 631
QY 8 FGQGYVQTPFLSESNVRYKISAGSCLPLTAGPSYVKFQDNPVGSQTF-SAGLHLR-V 64
Db 632 FAPGVSP 638
QY 65 FDPSTGA 71
```

```
RESULT 4
ENTRY A61046 #type complete
TITLE ecysone-induced membrane protein IMP-E3 - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 16-Feb-1997
ACCESSIONS A61046
REFERENCE A61046
#authors Moore, J.T.; Fristrom, D.; Hammonds, A.S.; Fristrom, J.W.
#journal Dev. Genet. (1990) 11:299-309
#title Characterization of IMP-E3, a gene active during imaginal disc morphogenesis in Drosophila melanogaster.
#accession A61046
##status preliminary
##molecule_type mRNA
##residues 1-331 ##label MOO
GENETICS
#gene FlyBase:ImpE3
##cross-references FlyBase:FBgn0001255
KEYWORDS membrane protein
SUMMARY #length 331 #molecular-weight 36583 #checksum 8221
Query Match 15.4%; Score 85; DB 2; Length 331;
Best Local Similarity 31.9%; Pred. No. 8.12e-01;
Matches 15; Conservative 12; Mismatches 18; Indels 2; Gaps 2;
Db 190 LDNFLRLYDDNYGRAAFSGESAMDRWSTASIAGKKRPPTKP-YVDF 235
QY 1 MEKFMAEFGQGYVQTPFLSESNVRYKI-SIAGSCLPLTAGPSYVKF 46
RESULT 5
ENTRY C56617 #type complete
TITLE cfac protein precursor - Escherichia coli plasmid NTP113
ORGANISM #formal_name Escherichia coli
DATE 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 20-Mar-1998
ACCESSIONS C56617
REFERENCE A56617
#authors Joridi, B.J.; Wallshaw, G.A.; van der Zeijst, B.A.; Gaastra, W.
#journal DNA Seq. (1992) 2:257-263
#title The complete nucleotide sequence of region 1 of the CFA/I fimbrial operon of human enterotoxigenic Escherichia coli.
#cross-references MUID:92329981
#accession C56617
##status preliminary
##molecule_type DNA
##residues 1-869 ##label JOR
##cross-references GB:M55661; NID:g145507; PID:g145510
##experimental_source enterotoxigenic strain, CFA/I-ST plasmid NTP113
##note sequence extracted from NCBI backbone (NCBIN:108960, NCBIP:108971)
GENETICS
#gene cfac
#genome plasmid
SUMMARY #length 869 #molecular-weight 97830 #checksum 9755
Query Match 15.4%; Score 85; DB 2; Length 869;
Best Local Similarity 32.4%; Pred. No. 8.12e-01;
Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;
Db 158 AFIQSQTINLSDSGYKRLISGNSALGITDTSY 191
QY 11 GYVQTPFLSESNVRYK-ISIAGSCLPLTAGPSY 43
RESULT 6
ENTRY S17402 #type complete
TITLE parasporal crystal protein cryIIc - Bacillus thuringiensis
```

(TM)

Result	Query			ID	Description	Pred. No.
	No.	Score	Match			
1	92	16.7	299	G51248	hypothetical protein	7.69e-02
2	89	16.2	566	2	probable OMP leader	2.14e-01
3	86	15.6	1693	1	genome polyprotein -	5.84e-01
4	85	15.4	331	2	ecdysone-induced memb	8.12e-01
5	85	15.4	869	2	cfaC protein precursor	8.12e-01
6	84	15.2	622	2	parasporal crystal pr	1.13e+00
7	83	15.1	281	2	hypothetical protein	1.56e+00
8	83	15.1	611	2	hypothetical protein	1.56e+00
9	82	14.9	81	2	tyrosine kinase - hum	2.15e+00
10	82	14.9	449	2	hypothetical protein	2.15e+00
11	82	14.9	527	2	tyrosine kinase - hum	2.15e+00
12	81	14.7	216	2	tyrosine kinase - hum	2.15e+00
13	81	14.7	527	2	flagellar L-ring prot	2.96e+00
14	81	14.7	527	2	protein-tyrosine kina	2.96e+00
15	81	14.7	3011	2	Txk - mouse	2.96e+00
16	80	14.5	171	2	polyprotein precursor	2.96e+00
17	80	14.5	572	2	hypothetical protein	4.05e+00
18	80	14.5	802	2	GRESAG protein - Tryp	4.05e+00
19	80	14.5	954	2	NAB3 protein - yeast	4.05e+00
20	80	14.5	1659	2	glucan 1,4-alpha-gluc	4.05e+00
21	80	14.5	3011	1	vitellogenin precursor	4.05e+00
22	79	14.3	370	2	genome polyprotein -	4.05e+00
23	79	14.3	472	1	hypothetical protein	5.55e+00
24	79	14.3	472	1	vitamin D-binding pro	5.55e+00


```
DE Amino acid sequence of lepidoteran-active 8612 toxin.
KW 8612 toxin; PCR: primer; amplification; Bacillus thuringiensis; probe;
KW lepidoptera; pest; pesticide; Ostrinia nubilalis; Heliothis virescens;
KW Helicoverpa zea; hybridisation.
OS Bacillus thuringiensis.
PN WQ9840490-AI.
PD 17-SEP-1998.
PF 13-MAR-1998; US-05081.
PR 13-MAR-1997; US-040512.
PA (MYCO ) MYCOGEN CORP.
PI Muller-Cohn J, Narva KE, Schnepf HE;
DR WPI; 98-506734/43.
DR N-PSDB; V52612.
PT New insecticidal Bacillus thuringiensis toxins - useful for
PT controlling lepidopteran pests, especially Ostrinia nubilalis,
PT Heliothis virescens and Helicoverpa zea
PS Claim 5: Pages 36-38; 50pp; English.
CC This is the amino acid sequence of a novel Bacillus thuringiensis toxin
CC used in the method of the invention, to control lepidopteran pests.
CC The new toxins are useful as pesticides, especially for the control of
CC Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The
CC polynucleotide coding sequences are useful for recombinant expression
CC of the toxins and the primers, together with probes derived from the
CC new sequences, are useful for the identification and characterisation
CC of novel genes that encode pesticidal toxins.
SQ Sequence 633 AA;

Query Match 14.9%; Score 82; DB 35; Length 633;
Best Local Similarity 22.0%; Pred. No. 2.27e+01;
Matches 13; Conservative 21; Mismatches 23; Indels 2; Gaps 2;

Db 511 fisekfgnqdsrfednsntartylrgrngnsynlylrssignstlrvtngryvtas 569
QY 12 YVQTPFLSESNVRYKIS-TAGSCPLSTAGPSYVKE-QDNPVGSGTFSAGLHLRVDPDS 68

RESULT 12
ID R66995 standard; Protein: 3011 AA.
AC R66995;
DT 01-AUG-1995 (first entry)
DE Hepatitis C virus gene HC-J1/protein.
KW Hepatitis C virus; HCV gene HC-J1/protein; specific antibodies.
OS Hepatitis C virus.
PN J06284887-A.
PD 11-OCT-1994.
PF 10-DEC-1993; 345753.
PR 10-DEC-1992; JP-360705.
PA (IMMO ) IMMUNO JAPAN KK.
DR WPI; 94-362594/45.
DR N-PSDB; Q74770.
PT HCV genes and the corresponding proteins - used in the production
PT of anti-HCV antibodies and the detection of HCV infection
PS Claim 11: Pages 18-32; 35pp; Japanese.
CC Q74770 encodes R66995 the HC-J1/protein, the cDNA can be used in
CC the construction of an expression vector for the transformation
CC of a host cell. The host cell can then be used in the production
CC of proteins and peptides, useful in the preparation of monoclonal
CC and polyclonal HCV-specific antibodies.
SQ Sequence 3011 AA;

Query Match 14.7%; Score 81; DB 12; Length 3011;
Best Local Similarity 40.0%; Pred. No. 2.72e+01;
Matches 18; Conservative 10; Mismatches 13; Indels 4; Gaps 4;

Db 2116 vpspefftdgvrllh-rfappckllreevsfrvghdyvpvgsq 2159
QY 13 VQTP-FLSESNVRYKISAGSC-PLSTAGPSY-VKFDNPNVGSQ 54

RESULT 13
ID R40120 standard; Protein: 3011 AA.
AC R40120;
DT 27-JAN-1994 (first entry)
```

```
DE HCV genomic amino acid sequence isolated from infected human LG.
KW Hepatitis C Virus; Non-A, non-B hepatitis Virus; HCV; NANBHV;
KW human growth hormone; HGH; secretion signal; fusion protein;
KW vaccine.
OS Hepatitis C Virus.
PN WQ9315193-A.
PD 05-AUG-1993.
PF 29-JAN-1993; U00907.
PR 31-JAN-1992; US-830024.
PA (ABBO ) ABBOTT LAB.
PI Bode SL, Casey JM, Desai SM, Devare SG, Fraill DE;
DR WPI; 93-258673/32.
DR 2116 vpspefftdgvrllh-rfappckllreevsfrvghdyvpvgsq 2159
PT New plasmid pHCV-162 is a mammalian expression systems for HCV
PT proteins - useful for diagnosing HCV infection and as vaccines
PT for preventing HCV infection
PS Example 1; Page 39-49; 100pp; English.
CC RNA was isolated from the plasma of a HCV seropositive human
CC (designated "LG") and cDNA was prepared from it. The cDNA was
CC PCR amplified using specific primers with sequences based
CC on the prototype HCV-1 cDNA sequence (GENBANK M62321). Further
CC amplification using nested primers resulted in 7 adjacent HCV DNA
CC fragments which could be assembled into a full-length sequence. The
CC DNA sequence was determined and translated into the genomic amino
CC acid sequence. Comparison of the LG genomic amino acid sequence
CC with that from HCV-1 showed 134 amino acid differences.
SQ Sequence 3011 AA;

Query Match 14.7%; Score 81; DB 8; Length 3011;
Best Local Similarity 40.0%; Pred. No. 2.72e+01;
Matches 18; Conservative 10; Mismatches 13; Indels 4; Gaps 4;

Db 2116 vpspefftdgvrllh-rfappckllreevsfrvghdyvpvgsq 2159
QY 13 VQTP-FLSESNVRYKISAGSC-PLSTAGPSY-VKFDNPNVGSQ 54

RESULT 14
ID P92033 standard; Protein: 170 AA
AC P92033
DT 02-MAR-1990 (first entry)
DE Sequence encoded in the hepatitis C virus (HCV) cDNA insert in clone 14c.
KW Hepatitis C virus (HCV); non-A, non-B hepatitis (NANBH).
OS Hepatitis C virus.
FH Key Location/Qualifiers
FT region 1..71
FT /*tag= a
PN EP-318216-A.
PD 31-MAY-1989.
PF 18-NOV-1988; 310922.
PR 14-NOV-1988; US-271450; US-122714.
PA (CHIR) Chiron Corp.
PI Houghton M, Choo q-L, Kuo G;
DR WPI; 89-159274/22.
DR N-PSDB; N92089.
PT Purified hepatitis C virus
PT - and associated nucleic acids and polypeptide(s)
PS Claim 13; Figure 18; 139pp; English.
CC It is the sequence encoded in the hepatitis C virus (HCV) cDNA insert in
CC clone 14c. Tag a - the region of overlap with the HCV antigen encoded in
CC clone 25c. It is an epitope which could be used as immunoassay reagents
CC and vaccines and to generate antibodies useful in diagnosis and passive
CC immunotherapy for HCV infection/non-A, non-B hepatitis.
SQ Sequence 170 AA;

Query Match 14.5%; Score 80; DB 1; Length 170;
Best Local Similarity 37.8%; Pred. No. 3.26e+01;
Matches 17; Conservative 11; Mismatches 13; Indels 4; Gaps 4;

Db 63 vpspefftdgvrllh-rfappckllreevsfrvghdyvpvgsq 106
QY 13 VQTP-FLSESNVRYKISAGSC-PLSTAGPSY-VKFDNPNVGSQ 54
```


AC W76368; 1998 (first entry)
 DT 03-DEC-1998 (first entry)
 DE Hepatitis E virus hollow particle protein #1.
 KW Hollow particle protein; virus; antibody; detection; immunoassay;
 OS Infection.
 QS Hepatitis virus.
 FH Key Location/Qualifiers
 FT Protein 1..1693 /note= "Partial sequence"
 FT J10234383-A.
 PN 08-SEP-1998.
 PD 28-FEB-1997; 062445.
 PR 28-FEB-1997; JP-062445.
 PA (DENK-) DENKA SEIKEN KK.
 PA (KOKU-) KOKURITSU YOBO EISEI KENKYUSHO.
 DR WPI: 98-535037/46.
 DR N-PSDB: V61687.
 DT Hepatitis E virus hollow particle poly:peptide(s) and nucleic acids
 PT encoding it - useful for more accurate detection of HEV in samples,
 PT using immuno-assays and nucleic acid hybridisation
 PS Claim 10; Page 17-24; 29pp; Japanese.
 CC This sequence represents a Hepatitis E viral hollow particle protein.
 CC This polypeptides can be used to raise antibodies to detect HEV
 CC infection in samples, e.g. by immuno-assay based techniques, and the
 CC nucleic acid can be used for the same in nucleic acid hybridisation
 CC assays. The polypeptides and nucleic acids allow more accurate
 CC detection of HEV than previously possible.
 SQ Sequence 1693 AA;
 Query Match 17.4%; Score 96; DB 35; Length 1693;
 Best Local Similarity 26.9%; Pred. No. 1.66e+00;
 Matches 18; Conservative 21; Mismatches 25; Indels 3; Gaps 3;
 Db 572 frtsfdvgavleangperynlfsdasqstmaagpfsptyaasaglevryvaagldhrav 631
 QY 8 FGQGYVQTPLSESNVRYKISAGSCPLSTAGSPSYVKFQDNPVGSOT-F-SAGLHLR-V 64
 Db 632 fapgvsp 638
 QY 65 FDPSTGA 71
 RESULT 3
 ID W75773 standard; Protein; 623 AA.
 AC W75773;
 DT 02-DEC-1998 (first entry)
 DE Amino acid sequence of lepidoteran-active HD573 toxin.
 KW HD573 toxin; PCR; primer; amplification; Bacillus thuringiensis; probe;
 KW lepidoptera; pest; pesticide; Ostrinia nubilalis; Heliothis virescens;
 KW Helicoverpa zea; hybridisation.
 OS Bacillus thuringiensis.
 PN WO9840490-A1.
 PD 17-SEP-1998.
 PF 13-MAR-1998; U05081.
 PR 13-MAR-1997; US-040512.
 PA (MYCO) MYCOGEN CORP.
 PI Muller-Cohn J, Narva KE, Schnepf HE;
 DR WPI: 98-506734/43.
 DR N-PSDB: V52610.
 DT New insecticidal Bacillus thuringiensis toxins - useful for
 PT controlling lepidopteran pests, especially Ostrinia nubilalis,
 PT Heliothis virescens and Helicoverpa zea
 PS Claim 14; Pages 28-30; 50pp; English.
 CC This is the amino acid sequence of a novel Bacillus thuringiensis toxin
 CC used in the method of the invention, to control lepidopteran pests.
 CC The new toxins are useful as pesticides, especially for the control of
 CC Ostrinia nubilalis, Heliothis virescens, and Helicoverpa zea. The
 CC polynucleotide coding sequences are useful for recombinant expression
 CC of the toxins and the primers, together with probes derived from the
 CC new sequences, are useful for the identification and characterisation
 CC of novel genes that encode pesticidal toxins.
 SQ Sequence 623 AA;

Query Match 15.8%; Score 87; DB 35; Length 623;
 Best Local Similarity 18.2%; Pred. No. 9.05e+00;
 Matches 12; Conservative 26; Mismatches 26; Indels 2; Gaps 2;
 Db 502 fisekyngqdsirfseisntartylrgnsgnsynlylrsvsgststirvtingrvytanv 561
 QY 12 YVQTPFLSESNVRYKIS-IGASCPSTAGSPSYVKF-QDNPVGSQTF-SAGLHLR-VFDPST 69
 Db 562 nttnnn 567
 QY 70 GALVDS 75
 RESULT 4
 ID R51264 standard; Protein; 1693 AA.
 AC R51264;
 DT 21-OCT-1994 (first entry)
 DE HEV strain protein encoded by ORF-1.
 KW Hepatitis E virus; HEV; strain SAR-55; open reading frame; ORF;
 KW antibody; detection; diagnosis; primates; stool suspension.
 OS Hepatitis E virus strain SAR-55.
 PN WO9406913-A.
 PD 31-MAR-1994.
 PF 17-SEP-1993; U08849.
 PR 18-SEP-1992; US-947263.
 PA (USSH) US SEC DEPT HEALTH.
 PI Emerson SU, Purcell RH, Tsarev SA;
 DR WPI: 94-118462/14.
 DR N-PSDB: Q45197.
 DT Purified hepatitis E strain SAR-55 virus - used to develop prods.
 PT for use in detection, diagnosis, vaccines and therapy of
 PT hepatitis E virus infection
 PS Disclosure: Page 57-62; 114pp; English.
 CC The sequences given in R51264-66 are encoded by the hepatitis E virus
 CC (HEV) strain SAR-55. The cDNA sequence contains three open reading
 CC frames (ORFs). These proteins can be used to stimulate the production
 CC of protective antibodies upon injection into a mammal that would serve
 CC to protect the mammal upon challenge with wild type HEV. The proteins
 CC can be used for detection and diagnosis of HEV infection. The HEV
 CC SAR-55 cDNA was isolated from primates inoculated with stool
 CC suspensions obtained from hepatitis E patients.
 SQ Sequence 1693 AA;
 Query Match 15.6%; Score 86; DB 10; Length 1693;
 Best Local Similarity 25.4%; Pred. No. 1.09e+01;
 Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;
 Db 572 frtsfdvgavleangperhnlfsdasqstmaagpfsityaasaglevryvaagldhrav 631
 QY 8 FGQGYVQTPLSESNVRYKISAGSCPLSTAGSPSYVKFQDNPVGSOT-F-SAGLHLR-V 64
 Db 632 fapgvsp 638
 QY 65 FDPSTGA 71
 RESULT 5
 ID W80196 standard; Protein; 1693 AA.
 AC W80196;
 DT 23-DEC-1998 (first entry)
 DE Protein encoded by ORF1 of ET-NANB (HEV) Burma strain DNA sequence.
 KW Enterically transmitted non A non B hepatitis virus; ET-NANB;
 KW Hepatitis E virus; HEV; Burma HEV isolate; vaccine;
 KW diagnostic probe
 OS Non A non B Hepatitis virus.
 PN US5824649-A.
 PD 20-OCT-1998.
 PF 07-JUN-1995; 475807.
 PR 25-JUL-1994; US-279823.
 PR 17-JUN-1988; US-208997.
 PR 11-APR-1989; US-336672.
 PR 16-JUN-1989; US-367486.
 PR 13-OCT-1989; US-420921.

Search completed: Wed Sep 1 16:10:34 1999
Job time : 24 secs.

[illegible]

FT NON_TER **I**
FT NON_TER **275** **275**
SQ SEQUENCE **275 AA;** **30495 MW; 9BEIEE8B CRC32:**

Query Match 19.9%; Score 80; DB 14; Length 275;
Best Local Similarity 37.8%; Pred. No. 1.50e+00;
Matches 17; Conservative 11; Mismatches 13; Indels 4; Gaps 4

Db **140 VSPPEFFTELDGRLH-RFAPPCKPLLRREEVSFRVGLHEYPVGSG 183**
I : I :: I :: I : I
| : | :: | :: | : |
Oy **13 VOTP-FLSESNSVRKISIAGSC-PLSTAGPSY-VKFQDPNPGVG 54**
I : I :: I :: I : I
| : | :: | :: | : |

RESULT 5

ID O01993 PRELIMINARY; PRT; 357 AA.
AC O01993;
DT 01-JUL-1997 (TREMBREL. 04, CREATED)
DT 01-JUL-1997 (TREMBREL. 04, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBREL. 09, LAST ANNOTATION UPDATE)
DE VC27A/L.1 PROTEIN.
GN VC27A/L.1.
OS CAENORHADITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHADIDTIDA:
OC RHABDITINA; RHABDITOIDEA; RHABDITIACE; PELODERINAE; CAENORHADITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA WHITE S.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
Rn [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIESTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., McMURRAY A., NORTHMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D.; SHOMKEEN R.,
RA SHALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
DR ENBL; 295621; E1349927; .
SQ SEQUENCE 357 AA; 41307 MW; D2FB4683 CRC32;

Query Match 19.98%; Score 80; DB 5; Length 357;
Best Local Similarity 45.8%; Pred. No. 1.50e+00;
Matches 11; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Db **211 RSYLETFPYRDYNDSVKVPILGS 234**
I : I :: I :: I : I
| : | :: | :: | : |
Oy **10 QGVOTPFLESNSVRKIAGS 33**
I : I :: I :: I : I
| : | :: | :: | : |

RESULT 6

ID Q07034 PRELIMINARY; PRT; 802 AA.
AC Q07034;
DT 01-NOV-1996 (TREMBREL. 01, CREATED)
DT 01-NOV-1996 (TREMBREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBREL. 09, LAST ANNOTATION UPDATE)
DE RNA BINDING PROTEIN.
OS SACCHAROMYCETES CERVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMITASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RA SUGIMOTO K., MATSUMOTO K., KORNBURG R.D., REED S.I., WITTENBERG C.;
RT "HMd1 of the yeast Saccharomyces cerevisiae encodes a putative
RNA/single strand DNA-binding protein, whose overexpression reduces
the expression of a Gl cyclin, CLN3.";
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.

[W][O][R][L][D]

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Sep 1 16:10:10 1999; MasPar time 7.61 Seconds
401.470 Million cell updates/sec
Tabular output not generated.

Title: >PCT-US99-13024-2
Description: (1-56) from PCTUS9913024.pep (4 of 12)
Perfect Score: 402
Sequence: 1 MEXFMAEFGGQVQTPFLE.....STAGPSYVKFQDNPVGSQTF 56

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spiremb19

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organella
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 34.913; Variance 62.861; scale 0.555

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	81	20.1	3011	14	GENOME POLYPROTEIN	1.08e+00
2	80	19.9	275	14	POLYPROTEIN (FRAGMENT)	1.50e+00
3	80	19.9	275	14	POLYPROTEIN (FRAGMENT)	1.50e+00
4	80	19.9	275	14	POLYPROTEIN (FRAGMENT)	1.50e+00
5	80	19.9	357	5	VC27A7L.1 PROTEIN.	1.50e+00
6	80	19.9	802	3	RNA BINDING PROTEIN.	1.50e+00
7	80	19.9	2436	14	POLYPROTEIN (FRAGMENT)	1.50e+00
8	79	19.7	542	5	SIMILAR TO GLUTAMATE D	2.10e+00
9	79	19.7	860	1	360AA LONG HYPOTHETICA	2.10e+00
10	78	19.4	70	2	POLYPROTEIN.	2.91e+00
11	78	19.4	3011	14	POLYPROTEIN.	2.91e+00
12	78	19.4	3011	14	POLYPROTEIN.	2.91e+00
13	78	19.4	3011	14	POLYPROTEIN.	2.91e+00
14	78	19.4	3011	14	POLYPROTEIN.	2.91e+00
15	77	19.2	418	2	LECITHINASE.	4.04e+00
16	77	19.2	470	2	LECITHINASE.	4.04e+00
17	77	19.2	566	2	HYPOTHETICAL 63.5 KD P	4.04e+00
18	77	19.2	967	2	PEPB.	4.04e+00
19	77	19.2	1464	11	N-METHYL-D-ASPARTATE R	4.04e+00
20	77	19.2	1464	11	N-METHYL-D-ASPARTATE R	4.04e+00

21	77	19.2	1454	4	Q12879	N-METHYL-D-ASPARTATE R	4.04e+00
22	77	19.2	2219	5	Q23388	2K1067.2 PROTEIN.	4.04e+00
23	76	18.9	535	5	P91315	SIMILARITY TO HUMAN GT	5.58e+00
24	76	18.9	550	5	O02490	FRUCTOFURANOSIDASE (EC	5.58e+00
25	76	18.9	821	2	O51735	OUTER MEMBRANE PROTEIN	5.58e+00
26	76	18.9	1142	4	O14497	B120.	5.58e+00
27	76	18.9	2660	5	Q19785	F25F2.2 (FRAGMENT).	5.58e+00
28	74	18.4	364	3	O74624	MANNOSE-1-PHOSPHATE GU	7.70e+00
29	74	18.4	178	3	O60074	HYPOTHETICAL 19.4 KD P	1.06e+01
30	74	18.4	921	5	P90770	C34B7.2 PROTEIN.	1.06e+01
31	73	18.2	92	4	O72154	L1 PROTEIN (FRAGMENT).	1.45e+01
32	73	18.2	442	4	O43411	HYPOTHETICAL 49.3 KD P	1.45e+01
33	73	18.2	586	9	Q38005	TAIL PROTEIN.	1.45e+01
34	73	18.2	964	2	O53784	PUTATIVE MEMBRANE PROT	1.45e+01
35	73	18.2	1693	14	O81876	COMPLETE GENOME SEQUEN	1.45e+01
36	73	18.2	2301	10	O80784	F13P17.19 PROTEIN.	1.45e+01
37	72	17.9	347	14	O82650	HAEMAGGLUTININ (FRAGME	1.98e+01
38	72	17.9	347	14	O82658	HAEMAGGLUTININ (FRAGME	1.98e+01
39	72	17.9	347	14	O82649	HAEMAGGLUTININ (FRAGME	1.98e+01
40	72	17.9	362	14	O67383	HEMAGGLUTININ HAI DOMA	1.98e+01
41	72	17.9	362	14	O67381	HEMAGGLUTININ (FRAGMEN	1.98e+01
42	72	17.9	379	14	O67363	HEMAGGLUTININ (FRAGMEN	1.98e+01
43	72	17.9	379	14	O67366	HEMAGGLUTININ (FRAGMEN	1.98e+01
44	72	17.9	447	14	O89483	POLYPROTEIN (FRAGMENT)	1.98e+01
45	72	17.9	447	14	O89496	POLYPROTEIN (FRAGMENT)	1.98e+01

ALIGNMENTS

RESULT 1
ID Q03463; PRELIMINARY; PRT; 3011 AA.
AC Q03463;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN.
OS HEPATITIS C VIRUS (ISOLATE HC-J1) (HCV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE;
NC HEPATITIS C-LIKE VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93117120.
RA OKAMOTO H., KANAI N., MISHIRO S.;
RT "Full-length nucleotide sequence of a Japanese hepatitis C virus isolate (HC-J1) with high homology to USA isolates."
RL NUCLEIC ACIDS RES. 20:6410-6410(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91013116.
RA OKAMOTO H., OKADA S., SUGIYAMA Y., YOTSUMOTO S., TANAKA T.,
RA YOSHIZAWA H., TSUDA F., MIYAKAWA Y., MAYUMI M.;
RT "The 5'-terminal sequence of the hepatitis C virus genome."
RL JPN. J. EXP. MED. 60:167-177(1990).
DR EMBL; D10749; G221587; -
DR PFAM; PF00998; HCV_RGRP; 1.
DR PFAM; PF01001; HCV_NS4b; 1.
DR PFAM; PF01006; HCV_NS4a; 1.
KW POLYPROTEIN
SQ SEQUENCE 3011 AA; 327114 MW; 399EDBFA CRC32;
Query Match 20.1%; Score 81; DB 14; Length 3011;
Best Local Similarity 40.0%; Pred. No. 1.08e+00;
Matches 18; Conservative 10; Mismatches 13; Indels 4; Gaps 4;
Db 2116 VPSPEFFTELQVRLH-RFAPCKPLLEEVSRVGLHDYVGSQ 2159
QY 13 VQTP-FLSESNSVRYKISAGSC-PLSTAGPSY-VKFDNPVGSQ 54
RESULT 2
ID Q68468; PRELIMINARY; PRT; 275 AA.
AC Q68468;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

M P S R C H P P (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:17:15 1999; MasPar time 6.29 Seconds
314.294 Million cell updates/sec

Tabular output not generated.

Title: >PCT-US99-13024-2
Description: (1-93) from PCTUS9913024.pep (6 of 12)
Perfect Score: 634
Sequence: 1 MEKFMAEFGQGVQTPFLUSE.....DSKSYAFSTNDTTSAAFVS 93

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 27.254; Variance 116.070; scale 0.235

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Pred. No.	
Result No.	Score	Query Match	Length	ID	Description	
1	300	47.3	65	19	R97371	Phage T4 ORFX gene pr
2	96	15.1	1693	35	W76368	Hepatitis E virus hol
3	92	14.5	520	1	P94617	Neutral protease enco
4	92	14.5	521	3	P51009	Sequence of neutral p
5	87	13.7	242	19	P95000	Bacillus subtilis lic
6	87	13.7	623	35	W75773	Amino acid sequence o
7	86	13.6	239	1	R06621	Hybrid (1.3-1.4)-pre
8	86	13.6	1693	34	W71209	Protein encoded by OR
9	86	13.6	1693	19	R91813	Hepatitis E virus str
10	86	13.6	1693	10	R51264	HEV strain protein en
11	86	13.6	1693	37	W81519	Hepatitis E virus (HE
12	86	13.6	1693	36	W80196	Protein encoded by OR
13	85	13.4	633	35	W75774	Amino acid sequence o
14	84	13.2	1693	3	R14618	Protein encoded by OR
15	83	13.1	613	36	W73122	A. thaliana ethylene
16	83	13.1	613	15	R74632	QETR ethylene respons

17	82	12.9	633	35	W75775	Amino acid sequence o
18	81	12.8	345	10	R52824	GTP-cyclohydrolase II
19	81	12.8	3011	8	R40120	HCV genomic amino aci
20	81	12.8	3011	12	R66995	Hepatitis C virus gen
21	80	12.6	170	1	P90150	Sequence of hepatitis
22	80	12.6	411	16	R90934	HCV NS5 domain antige
23	80	12.6	499	38	W67010	HCV non-structural pr
24	80	12.6	509	2	R06519	Microspore-specific c
25	80	12.6	516	6	R33441	HCV NS5 encoded by pH
26	80	12.6	516	6	R33595	HCV CKS-NS5E recombin
27	80	12.6	516	7	R33633	HCV CKS-NS5E fusion p
28	80	12.6	516	4	R21566	HCV CKS-NS5E - pHCV-4
29	80	12.6	1766	1	P92041	Sequence encoded in t
30	80	12.6	1786	1	P90158	Protein sequence of h
31	80	12.6	2301	1	P92047	Sequence encoded in t
32	80	12.6	2435	5	R25135	HCV polypeptide 1.
33	80	12.6	2436	5	R28582	HCV amino acid sequen
34	80	12.6	2436	1	P92050	Sequence encoded in t
35	80	12.6	2462	1	P90288	Peptide encoded by co
36	80	12.6	2772	2	R08123	Hepatitis C virus pol
37	80	12.6	2816	7	R34009	HCV-1 polypeptid.
38	80	12.6	2894	13	R70230	Composite hepatitis C
39	80	12.6	2894	5	R24440	Composite HCV HC-J1/C
40	80	12.6	2955	2	R08124	Hepatitis C virus put
41	80	12.6	3011	4	R21519	Compiled HCV sequence
42	80	12.6	3011	6	R31621	Hepatitis C virus (HC
43	80	12.6	3011	26	W34480	HCV polypeptid.
44	80	12.6	3011	16	R90931	Hepatitis C virus pol
45	80	12.6	3011	28	W40038	HCV polypeptid.

ALIGNMENTS

RESULT 1
ID R97371 standard; Protein; 65 AA.
AC R97371;
DE 07-JAN-1997 (first entry)
DT Phage T4 ORFX gene product. gp34.
KW Phage T4; tail fibre protein; nanotechnology; nano-structure;
KW filter; molecular sieve.
OS Bacteriophage T4.
PN W09611947-A1.
PD 25-APR-1996.
PF 13-OCT-1995; UI3023.
PR 13-OCT-1994; US-322760.
PA (GOLD/) GOLDBERG E B.
PI Goldberg EB;
DR WPI; 96-221942/22.
DR N-PSDB; T29053.
PT New proteins derived from T4 phage tail fibre proteins - that can self assemble into nano-structure(s), useful as filters etc, also corresponding DNA
PS Claim 7; Fig 7; 83pp; English.
CC A protein (R97371) of unspecified function is the product of open reading frame X of the tail fibre protein gene region (see also T29053) of phage T4. This gene region also includes open reading frames for tail fibre proteins (see also R97370 and R97372-74).
CC Tail fibre proteins (native or modified) can be produced in large quantities in microbial cells and used as building blocks of strong, stable nanostructures.
CC Sequence 65 AA;
Query Match 47.3%; Score 300; DB 19; Length 65;
Best Local Similarity 82.1%; Pred. No. 1.86e-18;
Matches 46; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
Db 1 mekfmaefggqvqtpflse...vrykistagscplstagsvkvkfdnpvgstf 56
QY 1 MEKFMAEFGQGVQTPFLSESNVRYKISITAGSCPLSTAGPSYVKFDNPVGSQTF 56
RESULT 2
ID W76368 standard; Protein; 1693 AA.

AC W76368;
 DT 03-DEC-1998 (first entry)
 DE Hepatitis E virus hollow particle protein #1.
 KW Hollow particle protein; virus; antibody; detection; immunoassay;
 OS infection.
 FH Hepatitis virus.
 FH Key Location/Qualifiers
 FT Protein 1..1693
 FT /note= "Partial sequence"
 PN J10234383-A.
 PD 08-SEP-1998.
 PF 28-FEB-1997; 062445.
 PR 28-FEB-1997; JP-062445.
 PA (DENK-) DENKA SEIKEN KK.
 PA (KOKU-) KOKURITSU YODO EISEI KENKYUSHO.
 DR WPI: 98-535037/46.
 DR N-PSDB: V61687.
 PT Hepatitis E virus hollow particle poly:peptide(s) and nucleic acids
 PT encoding it - useful for more accurate detection of HEV in samples,
 PT using immuno-assays and nucleic acid hybridisation
 PS Claim 10; Page 17-24; 29pp; Japanese.
 CC This sequence represents a Hepatitis E viral hollow particle protein.
 CC This polypeptides can be used to raise antibodies to detect HEV
 CC infection in samples, e.g. by immuno-assay based techniques, and the
 CC nucleic acid can be used for the same in nucleic acid hybridisation
 CC assays. The polypeptides and nucleic acids allow more accurate
 CC detection of HEV than previously possible.
 SQ Sequence 1693 AA;
 Query Match 15.1%; Score 96; DB 35; Length 1693;
 Best Local Similarity 26.9%; Pred. No. 3.01e+00;
 Matches 18; Conservative 21; Mismatches 25; Indels 3; Gaps 3;
 Db 572 frtsvdgavleangperynisfdaqstmaagpfspptyaasaaglevryvaagldhrav 631
 Qy 8 FGQGVQTFPLESSESVRYKTSIAGCPLSTAGPSYRVFQDNPVGSQF-F-SAGUHLR-V 64
 Db 632 fapgvsp 638
 Qy 65 FDPSTGA 71
 RESULT 3
 ID P94617 standard; protein: 520 AA.
 AC P94617;
 DT 21-JUN-1990 (first entry)
 DE Neutral protease encoded by npr gene.
 KW Protease; expression systems; subtilin; neutral protease; ds.
 OS Bacillus amyloliquefaciens.
 FH Key Location/Qualifiers
 FT protein 221..520
 FT /note="Mature"
 FT region 27..520
 FT /note="PRO-"
 PN US4801537-A.
 PD 31-JAN-1989.
 PF 29-MAR-1985; 717800.
 PR 29-MAR-1985; US-717800.
 PA (GENE-) Genex Corp.
 PI Nagarajan V, Rhodes CS, Banner CDB;
 DR WPI: 89-053629/07.
 DR N-PSDB: N91114.
 PT vectors for expression of polypeptide(s) in Bacillus -
 PT contg. promoter and regulatory regions which control expression
 PT and secretion of protease(s) in Bacillus.
 PS Disclosure; p: English.
 CC Claimed replicon comprises a promoter and regulatory regions, capable of
 CC expressing alkaline and neutral protease genes.
 SQ Sequence 520 AA;
 Query Match 14.5%; Score 92; DB 1; Length 520;
 Best Local Similarity 27.8%; Pred. No. 6.20e+00;
 Matches 15; Conservative 19; Mismatches 18; Indels 2; Gaps 2;

Db 239 lssesgkylvldlskptgtqitydlqnreynlp-gtlvsstnqftssgraa 291
 Qy 36 LSTAGPSIV-KFQDNPVGSQTFSSAGLHLRVFDPSTGALVDSKSYAFSTSDTTS 88
 RESULT 4
 ID P51009 standard; Protein: 521 AA.
 AC P51009;
 DT 01-DEC-1991 (first entry)
 DE Sequence of neutral protease encoded by the npr[BamP] gene.
 KW Bacillus expression vector; secretion vector.
 OS Bacillus amyloliquefaciens.
 FH Key Location/Qualifiers
 FT peptide 1..27
 FT /label= signal
 FT peptide 28..221
 FT /label= Pro sequence
 FT protein 222..521
 FT misc_difference 1
 FT /label= fMet
 PN EP-133756-A.
 PD 06-MAR-1985.
 PF 06-JUL-1984; 304662.
 PR 06-JUL-1983; US-511198.
 PR 08-JUN-1984; US-618902.
 PR 29-MAR-1985; US-717800.
 PA (GENE-) GENEX CORP.
 PI Nagarajan V, Banner CDB, Rhodes CS;
 DR WPI: 85-057299/10.
 DR N-PSDB: N50542.
 PT Replicable plasmidic expression vector - for transformation of
 PT Bacillus to direct expression of poly:peptide
 PS Disclosure; Fig 4; 37pp; English.
 CC The inventors claim a vector comprising a replicable plasmid
 CC containing the promoter and regulatory region of a gene selected
 CC from apr[BamP] and npr[BamP], for transformation of Bacillus to
 CC direct expression of polypeptide.
 SQ Sequence 521 AA;
 Query Match 14.5%; Score 92; DB 3; Length 521;
 Best Local Similarity 27.8%; Pred. No. 6.20e+00;
 Matches 15; Conservative 19; Mismatches 18; Indels 2; Gaps 2;

Db 240 lssesgkylvldlskptgtqitydlqnreynlp-gtlvsstnqftssgraa 292
 Qy 36 LSTAGPSIV-KFQDNPVGSQTFSSAGLHLRVFDPSTGALVDSKSYAFSTSDTTS 88
 RESULT 5
 ID P95000 standard; Protein: 242 AA.
 AC P95000;
 DT 12-FEB-1997 (first entry)
 DE Bacillus subtilis lichenase.
 KW Beer production; fermentation; barley; beta-glucan; hydrolysis;
 OS lichenase.
 OS Bacillus subtilis Y-25.
 PN J01067181-A.
 PD 13-MAR-1989.
 PF 08-SEP-1987; 224615.
 PR 08-SEP-1987; JP-224615.
 PA (ASAK) ASAKI BREWERIES KK.
 DR WPI: 89-119863/16.
 DR N-PSDB: N95000.
 PT Recombinant plasmid used in beer prodn. - obtd. by integrating
 PT lichenase gene derived from Bacillus subtilis, into vector
 PS Disclosure; Fig 4; 7pp; Japanese.
 CC The lichenase gene from Bacillus subtilis Y-25 is used for
 CC transforming Bacillus hosts so that they show increased lichenase
 CC expression. The recombinant lichenase enzyme produced by the
 CC transformants is useful in beer production for decomposing beta-
 CC glucan from barley.
 SQ Sequence 242 AA;

PI Borries B. Hofemeister J, Thomsen KK, Olsen O, Vonwettstein D;
DR WPI: 90-275129/36.
DR N-PSDB: Q05832.
PT New thermotable (1,3-1,4)-beta-glucanase - prep'd. using hybrid
PT gene obt'd. using Bacillus amyloliquefaciens and B.macerans genes
PS Disclosure; page 26; 84pp; English.
CC This hybrid protein is encoded by the beta-glucanase-HI gene.
CC Following processing of the signal peptide the mature protein
CC is produced, comprising the amino terminus of the amylolique-
CC faciens beta-glucanase and the carboxyl-terminal half of the
CC B.macerans beta-glucanase. This hybrid protein is thermostable
CC and hydrolyses beta-glycosidic linkages in (1,3-1,4)-beta-glucans.
CC Reducing sugars are obt'd. at high temps. and thus this enzyme can
CC be used in the mfr. of food prods. esp. beer and animal feed (eg
CC for feeding poultry). See also Q05833.
SQ Sequence 239 AA;

Query Match 13.6%; Score 86; DB 1; Length 239;
Best Local Similarity 25.5%; Pred. No. 1.80e+01;
Matches 14; Conservative 20; Mismatches 18; Indels 3; Gaps

Db 74 laltspynkfcdgnsrvtygyglyevrm-kpakntgivsvfftygtptgtp 127
|: ::||| |: |::| |: |::| |: |::| |: |::| |: |::| |: |::| |: |::|
Qy 36 LSTAGSYVRFQ-DNPVGSQTFSGAGLH-LRVFDPSTGALVDSKSYAFSTSDTTS 88

RESULT 8

ID W12009 standard; Protein; 1693 AA.
AC W1209;
DT 30-OCT-1998 (first entry)
DE Protein encoded by ORF 1 of the Burmese isolate of ET-NANB.
KW Enterically transmitted nonA/nonB hepatitis virus; identification;
KW HEV; ET-NANB; detection; vaccine.
OS Hepatitis virus.
FH Key Location/Qualifiers
FT Misc_difference 154 /note= "not specified"
FT Misc_difference 1514 /note= "not specified"
FT Misc_difference 1552 /note= "not specified"
FT Misc_difference 1552 /note= "not specified"
PN US5789559-A.
PD 04-AUG-1998.
PR 25-JUN-1994; 279823.
PR 05-APR-1991; US-681078.
PR 17-JUN-1988; US-208997.
PR 11-APR-1989; US-336672.
PR 19-JUN-1989; US-367486.
PR 13-OCT-1989; US-420921.
PR 05-APR-1990; US-505888.
PR 25-JUL-1994; US-279823.
PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Bradley DW, Fry KE, Krawczynski KZ, Reyes GR, Tam A,
PI Harbough PO;
DR WPI: 98-446186/38.
DR N-PSDB: V54729.
PT Hepatitis E virus DNA - useful for e.g. virus detection and viral
PT protein production
PS Disclosure; Columns 55-64; 45pp; English.
CC W1209-11 represent the proteins encoded by the open reading frames
CC (ORFs) of the DNA sequence of the Burmese isolate of an enterically
CC transmitted nonA/nonB viral hepatitis agent (ET-NANB). The nucleic
CC acid transgene may be used for identifying and sequencing the entire
CC viral agent (also referred to as HEV), detecting ET-NANB in
CC infected samples, e.g. by specific amplification of virus-derived DNA
CC sequences and for producing recombinant viral proteins for use in
CC vaccines.
SQ Sequence 1693 AA;

Query Match 13.6%; Score 86; DB 34; Length 1693;
Best Local Similarity 25.4%; Pred. No. 1.80e+01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

msrch_pp      protein - protein database search, using Smith-Waterman algorithm
Run on:      Wed Sep 1 16:31:44 1999;      MasPar time 5.61 Seconds
Tabular output not generated.      126.558 Million cell updates/sec

```

```
>PCT-US99-13024-2
Title:
Description:
(81-93) from PCTUS9913024.pep (12 of 12)
Perfect Score:
68
Sequence:
1 STSNDTTSAAEVS 13
```

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

Database: sptrembl9

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 21.936; Variance 22.856; scale 0.960

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	54	79.4	101	11	088517	CALCIUM CHANNEL BETA 1	7.52e-02
2	54	79.4	579	4	015331	L-TYPE CALCIUM CHANNEL	7.52e-02
3	47	69.1	241	5	020870	SIMILAR TO GLUTAMATE R	4.20e+00
4	47	69.1	982	5	Q09532	HYPOTHETICAL 110.9 KD	4.20e+00
5	47	69.1	2100	3	P87112	PUTATIVE TRANSCRIPTION	4.20e+00
6	46	67.6	268	5	Q94484	ORF DG1007 (FRAGMENT)	7.24e+00
7	45	66.2	337	5	F19474	F15A2.4 PROTEIN.	1.24e+01
8	45	66.2	583	3	Q12372	CHROMOSOME XII READING	1.24e+01
9	44	64.7	37	5	Q26818	SURFACE ANTIGEN (FRAG	2.09e+01
10	44	64.7	53	5	Q23592	SIMILAR TO ADENOSINE A	2.09e+01
11	44	64.7	236	11	Q00090	POTASSIUM CHANNEL PROT	2.09e+01
12	44	64.7	371	10	Q05584	HYPOTHETICAL 40.8 KD P	2.09e+01
13	44	64.7	461	2	Q68923	GLYCERALDEHYDE-3-PHOS	2.09e+01
14	44	64.7	488	5	Q44317	REVERSE TRANSCRIPTASE	2.09e+01
15	44	64.7	545	3	P87306	HYPOTHETICAL 62.5 KD P	2.09e+01
16	44	64.7	575	5	Q26040	SHAL 1 POTASSIUM CHANN	2.09e+01
17	44	64.7	611	11	Q3286	VOLTAGE-GATED K+ CHANN	2.09e+01
18	44	64.7	637	5	Q16932	W03F9.4 PROTEIN.	2.09e+01
19	44	64.7	636	11	P70632	KV4.3 POTASSIUM CHANNE	2.09e+01
20	44	64.7	636	4	Q60576	KV4.3 POTASSIUM CHANNE	2.09e+01

```
RP SEQUENCE FROM N.A.
RA POWERS P.A., GREGG R.G., HOGAN K.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U86953; G2155256; JOINED.
DR EMBL: U86954; G2155256; JOINED.
DR EMBL: U86955; G2155256; JOINED.
DR EMBL: U86956; G2155256; JOINED.
DR EMBL: U86957; G2155256; JOINED.
DR EMBL: U86958; G2155256; JOINED.
DR PFAM: PF00018; SH3; 1.
DR PFAM: PF00774; Ca_channel_B; 1.
SQ SEQUENCE 579 AA; 63705 MW; 651614EF CRC32;

Query Match 79.4%; Score 54; DB 4; Length 579;
Best Local Similarity 75.0%; Pred. No. 7.52e-02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTSNSFV 55
QY 81 STSNDTTSAAFV 92

RESULT 3
ID Q20870; PRELIMINARY; PRT; 241 AA.
AC Q20870;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SIMILAR TO GLUTAMATE RECEPTORS.
GN F56D1.7.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
RA CRAXTON A., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.; III of C.
RT "2.2 kb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA CHISSE S., WILSON R.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U39997; G1055122; -.
SQ SEQUENCE 241 AA; 27445 MW; F0977197 CRC32;

Query Match 69.1%; Score 47; DB 5; Length 241;
Best Local Similarity 80.0%; Pred. No. 4.20e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 59 GNDTISAADFV 68
QY 83 SNDTTSAAFV 92

RP SEQUENCE FROM N.A.
RA POWERS P.A., GREGG R.G., HOGAN K.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U86953; G2155256; JOINED.
DR EMBL: U86954; G2155256; JOINED.
DR EMBL: U86955; G2155256; JOINED.
DR EMBL: U86956; G2155256; JOINED.
DR EMBL: U86957; G2155256; JOINED.
DR EMBL: U86958; G2155256; JOINED.
DR PFAM: PF00018; SH3; 1.
DR PFAM: PF00774; Ca_channel_B; 1.
SQ SEQUENCE 579 AA; 63705 MW; 651614EF CRC32;

Query Match 79.4%; Score 54; DB 4; Length 579;
Best Local Similarity 75.0%; Pred. No. 7.52e-02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTSNSFV 55
QY 81 STSNDTTSAAFV 92

RESULT 4
ID Q09532; PRELIMINARY; PRT; 982 AA.
AC Q09532;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 110.9 KD PROTEIN F07F6.6 IN CHROMOSOME II.
GN F07F6.6.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIDA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA CHISSE S.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: TO GLUTAMATE RECEPTOR.
DR EMBL: U23486; G746452; -.
DR WORMPEP; F07F6.6; CE01898.
DR PFAM: PF00060; lig_chan; 1.
KW HYPOTHETICAL PROTEIN.
FT DOMAIN 827 830 POLY-ARG.
SQ SEQUENCE 982 AA; 110892 MW; B4285770 CRC32;

Query Match 69.1%; Score 47; DB 5; Length 982;
Best Local Similarity 80.0%; Pred. No. 4.20e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 59 GNDTISAADFV 68
QY 83 SNDTTSAAFV 92

RESULT 5
ID P87112; PRELIMINARY; PRT; 2100 AA.
AC P87112;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE PUTATIVE TRANSCRIPTIONAL REGULATOR C20G8.06.
GN SPAC20G8.06.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-972;
RA BADCOCK K., CHURCHER C.M., WOOD V., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: TO YEAST NOT1.
DR EMBL: Z95334; E315491; -.
KW HYPOTHETICAL PROTEIN; NUCLEAR PROTEIN; TRANSCRIPTION REGULATION.
SQ SEQUENCE 2100 AA; 237132 MW; 9A710DB4 CRC32;

Query Match 69.1%; Score 47; DB 3; Length 2100;
Best Local Similarity 66.7%; Pred. No. 4.20e+00;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1601 SATNDETKAAVF 1612
QY 81 STSNDTTSAAFV 92

RESULT 6
ID Q94484; PRELIMINARY; PRT; 268 AA.
AC Q94484;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)
DE ORF DGL007 (FRAGMENT).
OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
```


OC EUKARYOTA; DICTYOSTELIIDA; DICTYOSTELIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AX4;
 RA LOOMIS W.F.;
 RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: U56911; G1519534;
 FT NON_TER 268
 SQ SEQUENCE 268 AA; 30503 MW; 04E723F5 CRC32;
 Query Match 67.6%; Score 46; DB 5; Length 268;
 Best Local Similarity 53.8%; Pred. No. 7.24e+00;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Db 15 TTNNTTPNKEVS 27
 QY 81 STSNDTTSAAVFS 93
 RESULT 7
 ID Q19474 PRELIMINARY; PRT; 337 AA.
 AC Q19474;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE F15A2.4 PROTEIN.
 GN F15A2.4.
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEAE; RHABDITIDA; RHABDITIDA;
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA GREGORY J.;
 RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94150718.
 RA WILSON R., AINSKOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SHALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RL NATURE 368:32-38(1994).
 DR EMBL: 270207; E1345607;
 SQ SEQUENCE 337 AA; 38483 MW; ABD004B8 CRC32;
 Query Match 66.2%; Score 45; DB 5; Length 337;
 Best Local Similarity 56.7%; Pred. No. 1.24e+01;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Db 317 SKHNTTSATFV 328
 QY 81 STSNDTTSAAVFS 92
 RESULT 8
 ID Q12372 PRELIMINARY; PRT; 583 AA.
 AC Q12372;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE CHROMOSOME XII READING FRAME ORF YL061W.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
 OC SACCHAROMYCETACEAE; SACCHAROMYCES.
 RN [1]

RP SEQUENCE FROM N.A.
 RA WEDLER H., WEDLER E., SCHARFE M., WAMBUTT R.;
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MIPS;
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RA WEDLER H., WAMBUTT R.;
 RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: Z73166; E245763;
 DR EMBL: Z47973; G642323;
 DR PFAM: PF00324; aa_permeases; 1.
 SQ SEQUENCE 583 AA; 64217 MW; BDE5153C CRC32;
 Query Match 66.2%; Score 45; DB 3; Length 583;
 Best Local Similarity 50.0%; Pred. No. 1.24e+01;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
 Db 330 SSEDITASPVF 341
 QY 81 STSNDTTSAAVFS 92
 RESULT 9
 ID Q26818 PRELIMINARY; PRT; 37 AA.
 AC Q26818;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE SURFACE ANTIGEN (FRAGMENT).
 OS TRYPAPOSOMA BRUCEI BRUCEI.
 OC EUKARYOTA; EULENZOEA; KINETOPLASTIDA; TRYPAPOSOMATIDAE; TRYPAPOSOMA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-GAMBIENSE;
 RX MEDLINE: 83169719.
 RA MERRITT S.C., TSCHUDI C., KONIGSBERG W.H., RICHARDS F.F.;
 RT *Reverse transcription of trypanosome variable antigen mRNAs
 RT initiated by a specific oligonucleotide primer.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 80:1536-1540(1983).
 DR EMBL: V01391; G10560;
 FT NON_TER 1
 SQ SEQUENCE 37 AA; 4068 MW; 551442D4 CRC32;
 Query Match 64.7%; Score 44; DB 5; Length 37;
 Best Local Similarity 63.6%; Pred. No. 2.09e+01;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Db 14 TSNTTASNFV 24
 QY 82 TSNDTTSAAVFS 92
 RESULT 10
 ID Q23582 PRELIMINARY; PRT; 53 AA.
 AC Q23582;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE SIMILAR TO ADENOSINE A3 RECEPTOR. NCBI GI: 1072232.
 GN ZK721.5.
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEAE; RHABDITIDA; RHABDITIDA;
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94150718.
 RA WILSON R., AINSKOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,

RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPOAT J., WOHLDMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL NATURE 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MILLER N.;
 RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA WATERSTON R.;
 RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: U40951; G107232;
 SQ SEQUENCE 53 AA; 5833 MW; C4DA122E CRC32;

Query Match 64.7%; Score 44; DB 5; Length 53;
 Best Local Similarity 58.3%; Pred. No. 2.09e+01;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 6 STSNSTRVSFV 17
 ||||| : :||
 OY 81 STSNDTSAFVS 92

RESULT 11
 ID Q00090 PRELIMINARY; PRT; 236 AA.
 DT 01-NOV-1996 (TREMREL. 01, CREATED)
 DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
 DE POTASSIUM CHANNEL PROTEIN (FRAGMENT).
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
 OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA RUDY B., KENTROS C., DE MIERA E.;
 RL MOL. CELL. NEUROSCI. 2:89-102(1991).
 CC -!- FUNCTION: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT POTASSIUM
 ION PERMEABILITY OF EXCITABLE MEMBRANES.
 CC ASSUMING OPENED OR CLOSED CONFORMATIONS IN RESPONSE TO THE
 VOLTAGE DIFFERENCE ACROSS THE MEMBRANE, THE PROTEIN FORMS A
 POTASSIUM-SELECTIVE CHANNEL THROUGH WHICH K+ IONS MAY PASS IN
 ACCORDANCE WITH THEIR ELECTROCHEMICAL GRADIENT.
 CC MAY PLAY A ROLE IN THE NERVOUS SYSTEM AND IN THE REGULATION OF
 BEATING FREQUENCY IN PACEMAKER CELLS.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- THE SEGMENT S4 IS PROBABLY THE VOLTAGE-SENSOR AND IS
 CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT
 EVERY THIRD POSITION.
 CC -!- THE AMINO TERMINUS MAY BE IMPORTANT IN DETERMINING THE RATE OF
 INACTIVATION OF THE CHANNEL WHILE THE TAIL MAY PLAY A ROLE IN
 MODULATION OF CHANNEL ACTIVITY AND/OR TARGETING OF THE CHANNEL TO
 SPECIFIC SUBCELLULAR COMPARTMENTS.
 CC -!- SIMILARITY: THIS CHANNEL PROTEIN BELONGS TO THE A-TYPE POTASSIUM
 CURRENT CLASS.
 CC -!- SIMILARITY: BELONGING TO THE SHIV GENE SUBFAMILY, A HOMOLOGUE OF
 THE SHAKER GENE SHAL OF DROSOPHILA MELANOGASTER.
 DR EMBL: M74898; G205041;
 DR PFAM: PF00520; ion.trans. 1.
 KW IONIC CHANNEL; TRANSMEMBRANE; ION TRANSPORT; VOLTAGE-GATED CHANNEL;
 KW MULTIGENE FAMILY.
 FT TRANSMEM 15 33 SEGMENT S1 (BY SIMILARITY).
 FT TRANSMEM 57 78 SEGMENT S2 (BY SIMILARITY).
 FT TRANSMEM 89 110 SEGMENT S3 (BY SIMILARITY).
 FT TRANSMEM 118 136 SEGMENT S4 (BY SIMILARITY).

FT TRANSMEM 152 173 SEGMENT S5 (BY SIMILARITY).
 FT TRANSMEM 213 234 SEGMENT S6 (BY SIMILARITY).
 FT NON_TER 236 236
 SQ SEQUENCE 236 AA; 26261 MW; C4A458FB CRC32;

Query Match 64.7%; Score 44; DB 11; Length 236;
 Best Local Similarity 50.0%; Pred. No. 2.09e+01;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 108 TNNEDVSGAFVT 119
 |::: |::|:
 OY 82 TSNDTSAFVS 93

RESULT 12
 ID O65584 PRELIMINARY; PRT; 371 AA.
 AC O65584;
 DT 01-AUG-1998 (TREMREL. 07, CREATED)
 DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 40.8 KD PROTEIN.
 GN M389.70.
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; TRACHEOPHYTA;
 OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
 OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BEVAN M., VANDENBOL M., JALLET C., PORTELETTE D., HOHEISEL J.,
 RA MENES H.W., MAYER K., SCHUELLER C.;
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU ARABIDOPSIS SEQUENCING PROJECT;
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AL022223; E1283943;
 KW HYPOTHETICAL PROTEIN
 SQ SEQUENCE 371 AA; 40833 MW; 5B0F5AA1 CRC32;

Query Match 64.7%; Score 44; DB 10; Length 371;
 Best Local Similarity 38.5%; Pred. No. 2.09e+01;
 Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Db 224 SSSGSESSSFVS 236
 |::: |::|:
 OY 81 STSNDTSAFVS 93

RESULT 13
 ID O68923 PRELIMINARY; PRT; 461 AA.
 AC O68923;
 DT 01-AUG-1998 (TREMREL. 07, CREATED)
 DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
 DE GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE HOMOLOG.
 GN GAPX.
 OS STREPTOMYCES ROSEOFULVUS.
 OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
 OC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94237486.
 RA BIBB M.J., SHERMAN D.H., OMURA S., HOPWOOD D.A.;
 RT "Cloning, sequencing and deduced functions of a cluster of
 RT streptomycetes genes probably encoding biosynthesis of the polyketide
 RT antibiotic frenolicin.";
 RL GENE 142:31-39(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA REEVES C.D., SOLIDAY C.L.;
 RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF058302; G3170587;
 SQ SEQUENCE 461 AA; 49966 MW; 26EC2672 CRC32;

Query Match 64.7%; Score 44; DB 2; Length 461;
Best Local Similarity 46.2%; Pred. No. 2.09e+01;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
DB 394 TTSADAVSDFVC 406
QY 81 STSNDTTSAAFS 93

RESULT 14
ID 04317 PRELIMINARY; PRT; 488 AA.
AC 04317;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE REVERSE TRANSCRIPTASE (FRAGMENT).
GN R1 ORF.
OS DUGESIELLA SP.
OC EUKARYOTA; METAZOA; ARTHROPODA; CHELICERATA; ARACHNIDA; ARANEAE;
OC MYGALOMORPHAE; THERAPHOSIDAE; DUGESIELLA.
RN [1]
RP SEQUENCE FROM N.A.
RA BURKE W.D., MALIK H.S., EICKBUSH T.H.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF015813; G2736044; -.
KW RNA-DIRECTED DNA POLYMERASE.
FT NON_TER 1
SQ SEQUENCE 488 AA; 55092 MW; FA67B24C CRC32;

Query Match 64.7%; Score 44; DB 5; Length 488;
Best Local Similarity 58.3%; Pred. No. 2.09e+01;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
DB 251 SKTSDGTGAFFV 262
QY 81 STSNDTTSAAFS 92

RESULT 15
ID P87306 PRELIMINARY; PRT; 545 AA.
AC P87306;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 62.5 KD PROTEIN C31F10.04C IN CHROMOSOME II.
GN SPBG31F10.04C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA WOOD V., RAJANDREAM M.A., BARRELL B.G., POHL T.;
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; Z97204; E325252; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 545 AA; 62479 MW; 2F16D503 CRC32;

Query Match 64.7%; Score 44; DB 3; Length 545;
Best Local Similarity 54.5%; Pred. No. 2.09e+01;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
DB 364 SNEDTSTNEVT 374
QY 83 SNDTTSAAFS 93

Search completed: Wed Sep 1 16:32:18 1999
Job time : 34 secs.

```

SQ SEQUENCE 1036 AA; 109310 MW; C0DE5BE8 CRC32;
Query Match 70.6%; Score 48; DB 1; Length 1036;
Best Local Similarity 53.8%; Pred. No. 1.11e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 87 SSSNNNTSNGFVA 99
    :||: ||: ||:
QY 81 STSNDTTSAAVFS 93

RESULT 11
ID CPPLDROAC STANDARD; PRT; 532 AA.
AC 04220;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450 12B1 PRECURSOR (EC 1.14.-.-).
GN CYP12B1
OS DROSOPHILA ACANTHOPTERA (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A584.2;
RX MEDLINE; 97449781.
RA DANIELSON P.B., FOGLEMAN J.C.;
RT "Isolation and sequence analysis of cytochrome P450 12B1: the first
RT mitochondrial insect P450 with homology to 1 alpha,25 dihydroxy-D3
RT 24-hydroxylase."
RL INSECT BIOCHEM. MOL. BIOL. 27:595-604(1997).
CC -!- FUNCTION: PROBABLY INVOLVED IN STEROID HORMONES BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U78485; G2674278; -
CC DR FLYBASE; FBgn0022933; Dacc\Cyp12b1.
CC DR PROSITE; PS00086; CYTOCHROME P450; 1.
CC KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; HEME;
CC MW MITOCHONDRION; TRANSIT PEPTIDE.
CC FT TRANSIT 1 ? MITOCHONDRION (BY SIMILARITY).
CC FT CHAIN ? 532 CYTOCHROME P450 12B1.
CC FT BINDING 480 480 HEME (POTENTIAL).
CC SQ SEQUENCE 532 AA; 62031 MW; 53F97355 CRC32;
Query Match 67.6%; Score 46; DB 1; Length 532;
Best Local Similarity 77.8%; Pred. No. 3.44e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 335 DTTSAFVT 343
    ||||: ||:
QY 85 DTTSAAFVS 93

RESULT 12
ID YEM3_YEAST STANDARD; PRT; 1076 AA.
AC P40021;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHEICAL 119.3 KD PROTEIN IN PIP1-GLN3 INTERGENIC REGION.
GN YER033C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;

```

```

OC RN SACCHAROMYCETACEAE; SACCHAROMYCETES.
RP [1]
RC SEQUENCE FROM N.A.
RA STRAIN-S288C / AB972;
RA DIERICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,
RA AVILES E., BERNO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,
RA CHUNG E., DUNCAN M., GUZMAN E., HARTZELL G., HUNICKE-SMITH S.,
RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,
RA MOSEDALE D., NAKAHARA K., NAMATH A., NORGREN R., OEFNER P., OH C.,
RA PETEL F.X., ROBERTS D., SEHL P., SCHRAMM S., SHOGREN T., SMITH V.,
RA TAYLOR P., WEI Y., YELTON M., BOTSTEIN D., DAVIS R.W.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U18796; G603266; -
CC DR HYPOTHETICAL PROTEIN.
KW SEQUENCE 1076 AA; 119349 MW; 9FED783D CRC32;
Query Match 67.6%; Score 46; DB 1; Length 1076;
Best Local Similarity 50.0%; Pred. No. 3.44e+00;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 97 TSTNQTTNSFV 108
    :||: ||: ||:
QY 81 STSNDTTSAAVFS 92

RESULT 13
ID PURA_METJJA STANDARD; PRT; 345 AA.
AC Q57981;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ADENYLOSUCCINATE SYNTHETASE (EC 6.3.4.4) (IMP--ASPARTATE LIGASE).
GN PURA OR MJ0561.
OS METHANOCOCCUS JANNASCHII.
OC ARCHAEA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE;
OC METHANOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE; 96337999.
RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
RA OVERBEER R., KIRNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL SCIENCE 273:1058-1073(1996).
CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE DE NOVO PATHWAY OF PURINE
CC NUCLEOTIDE BIOSYNTHESIS.
CC -!- CATALYTIC ACTIVITY: GTP + IMP + L-ASPARTATE = GDP +
CC ORTHOPHOSPHATE + ADENYLOSUCCINATE.
CC -!- PATHWAY: FIRST COMMITTED STEP IN AMP BIOSYNTHESIS.
CC -!- SIMILARITY: WITH OTHER ADENYLOSUCCINATE SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; U67505; G1591267; -.
DR TIGR; MJ0561;
DR PROSITE; PS01266; ADENYLOSUCIN_SYN_1; 1.
DR PROSITE; PS00513; ADENYLOSUCIN_SYN_2; FALSE_NEG.
DR PFAM; PF00709; Adenylsucc_syt; 1.
DR HSP; P12283; IADI.
KW PURINE BIOSYNTHESIS; LIGASE; GTP-BINDING.
FT NP_BIND 18 345 AA; 37820 MW; E15BE2C9 CRC32;
SQ SEQUENCE 345 AA; 37820 MW; 8289EB89 CRC32;

Query Match 66.2%; Score 45; DB 1; Length 345;
Best Local Similarity 50.0%; Pred. No. 5.97e+00;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 200 TSKDTTASSFAA 211
QY 82 TSDNTTSAAFVS 93
||||:||||:

RESULT 14
ID IRL2_HCMVA STANDARD; PRT; 416 AA.
AC P16810;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DE 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN IRL12 (TRL12).
OS HUMAN CYTOMEGALOVIRUS (STRAIN AD169).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC BETAHERPESVIRINAE; CYTOMEGALOVIRUS.
CC [1]
RN SEQUENCE FROM N.A.
RP MEDLINE; 90269039.
RX CHEE M.S., BANKIER A.T., BECK S., BOHNI R., BROWN C.M., CERNY R.,
RA HORSNELL T., HUTCHISON C.A. III, KOUZARIDES T., MARIIGNETTI J.A.,
RA PREDDIE E., SACHWELL S.C., TOMLINSON P., WESTON K.M., BARRELL B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RL cytomegalovirus strain AD169."
CC -1- SIMILARITY: BELONGS TO RL11 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; X17403; E298604; -.
DR EMBL; X17403; E298631; -.
DR PIR; S09761; S09761.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 416 AA; 47416 MW; 8289EB89 CRC32;

Query Match 66.2%; Score 45; DB 1; Length 416;
Best Local Similarity 50.0%; Pred. No. 5.97e+00;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 42 TSPNTASTTFTV 53
QY 82 TSDNTTSAAFVS 93
||||:||||:

RESULT 15
ID POLG_BCMVN STANDARD; PRT; 3066 AA.
AC O65329;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN [CONTAINS: N-TERMINAL PROTEIN (P1); HELPER
DE COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KD PROTEIN

```

```

DE 1 (6K1); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KD PROTEIN 2 (6K2);
DE GENOME-LINKED PROTEIN (VPG); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA)
DE (EC 3.4.22.-) (49 KD PROTEINASE) (49 KD-PRO); NUCLEAR INCLUSION
DE PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);
DE COAT PROTEIN (CP)].
OS BEAN COMMON MOSAIC VIRUS (STRAIN NL-3 / MICHIGAN) (BCM).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
OC POTYVIRUS.
RN SEQUENCE FROM N.A.
RP MEDLINE; 96191623.
RX FANG G.W., ALLISON R.F., ZAMBOLIM E.M., MAXWELL D.P., GILBERTSON R.L.;
RT "The complete nucleotide sequence and genome organization of bean
RT common mosaic virus (NL3 strain).";
RL VIRUS RES. 39:13-23(1995).
CC -1- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID
CC TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.
CC -1- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
CC MAY BE INVOLVED IN REPLICATION.
CC -1- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.
CC -1- PTM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.
CC -1- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE
CC POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
CC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
CC INDIVIDUAL PROTEINS.
CC -1- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
CC -1- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
CC -1- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; U19287; G1373002; -.
DR PFAM; PF00271; helicase_C; 1.
DR PFAM; PF00580; RNA_dep_RNA_pol; 1.
DR PFAM; PF00767; Poty_coat; 1.
DR PFAM; PF00851; Peptidase_C6; 1.
DR PFAM; PF00863; Peptidase_C4; 1.
DR PFAM; PF00851; Peptidase_C6; 1.
KW HYDROLASE; TRANSFERASE; THIOL PROTEASE; RNA-DIRECTED RNA POLYMERASE;
KW COAT PROTEIN; POLYPROTEIN; COVALENT PROTEIN-RNA LINKAGE; HELICASE;
KW ATP-BINDING.
FT CHAIN 1 ? N-TERMINAL PROTEIN.
FT CHAIN ? ? HELPER COMPONENT PROTEINASE.
FT CHAIN ? ? PROTEIN P3.
FT CHAIN ? ? 6 KD PROTEIN 1.
FT CHAIN ? ? CYTOPLASMIC INCLUSION PROTEIN.
FT CHAIN ? ? 6 KD PROTEIN 2.
FT CHAIN ? ? GENOME-LINKED PROTEIN.
FT CHAIN ? ? NUCLEAR INCLUSION PROTEIN A.
FT CHAIN ? ? NUCLEAR INCLUSION PROTEIN B.
FT CHAIN ? ? COAT PROTEIN.
FT NP_BIND 1258 1265 ATP (POTENTIAL).
SQ SEQUENCE 3066 AA; 350387 MW; 67B875D6 CRC32;

Query Match 66.2%; Score 45; DB 1; Length 3066;
Best Local Similarity 54.5%; Pred. No. 5.97e+00;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 1194 STSKESTSTSF 1204
QY 81 STSNDTTSAAF 91
||||:||||:

Search completed: Wed Sep 1 16:31:25 1999
Job time : 11 secs.

```

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	100	18.1	1693	14	Q81876	COMPLETE GENOME SEQUEN	4.78e-03
2	96	17.4	114	14	Q41088	AC068L PROTEIN.	2.00e-02
3	92	16.7	299	14	Q65574	HYPOTHETICAL 32.4 KD P	8.15e-02
4	92	16.7	300	14	Q65581	UL7 POLYPEPTIDE.	8.15e-02
5	89	16.2	566	2	Q84354	HYPOTHETICAL 63.5 KD P	2.28e-01
6	86	15.6	1693	14	Q81862	ORF 1.	6.27e-01
7	86	15.6	1693	14	Q81344	HEPATITIS E VIRUS COM	6.27e-01
8	86	15.6	1693	14	Q89444	UNNAMED PROTEIN PRODUC	6.27e-01
9	86	15.6	1693	14	Q89440	METHYL TRANSFERASE.	6.27e-01
10	84	15.2	822	2	Q43743	CYTIC DELTA-ENDOTOXIN	1.21e+00
11	84	15.2	921	5	P90770	C34E7.2 PROTEIN.	1.21e+00
12	83	15.1	180	10	Q49984	PUTATIVE ETHYLENE RECE	1.68e+00
13	83	15.1	281	9	Q38022	ORF 1.	1.68e+00
14	83	15.1	1577	2	Q54178	GLUCOSYLTRANSFERASE.	1.68e+00
15	83	15.1	1693	14	Q39221	NONSTRUCTURAL POLYPROT	1.68e+00
16	81	14.7	196	14	Q65847	PUTATIVE COAT PROTEIN.	3.21e+00
17	81	14.7	216	2	Q67609	FLAGELLAR L-RING PROTE	3.21e+00
18	81	14.7	633	2	Q52764	INSECTICIDAL CRYSTAL P	3.21e+00
19	81	14.7	642	14	Q65848	PUTATIVE REATHTROUGH P	3.21e+00
20	81	14.7	3011	14	Q03463	GENOME POLYPROTEIN.	3.21e+00

```
RA LI Y., LU Z., SUN L., ROPP S., KUTISH G.F., ROCK D.L.,
RA VAN ETEN J.L.;
RT "Analysis of 74 Kb of DNA located at the right end of the 330-kb
RT chlorella virus PBCV-1 genome.";
RL VIROLOGY 237:360-377(1997).
DR EMBL: U42580; G2447164;
SQ SEQUENCE 114 AA; 13257 MW; 1B75BE43 CRC32;

Query Match 17.4%; Score 96; DB 14; Length 114;
Best Local Similarity 27.4%; Pred. No. 2,00e-02;
Matches 17; Conservative 18; Mismatches 25; Indels 2; Gaps 1;

Db 5 F5SM5NS1FVRLTASCAI--LNSYSGRKRNSLNFTWSLSLMVFVSIKFSMSIK 62
QY 17 FLSENSVRYKISAGSCPLSTAGPSYKFDQNPVGSQTFSSAGLHLRVDPSTGALVDSK 76
Db 63 KY 64
QY 77 SY 78

RESULT 3
ID Q65574 PRELIMINARY; PRT; 299 AA.
AC Q65574;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 32.4 KD PROTEIN.
GN UL7.
OS BOVINE HERPESVIRUS TYPE 1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.
RN
RP SEQUENCE FROM N.A.
RC STRAIN=COOPER, AND JURA;
RX MEDLINE; 95313343.
RA VLCEK C., BENES V., LU Z., KUTISH G.F., PAGES V., ROCK D.,
RA LETCHWORTH G.J., SCHWYZER M.;
RT "Nucleotide sequence analysis of a 30-kb region of the bovine
RT herpesvirus 1 genome which exhibits a colinear gene arrangement with
RT the UL21 to UL4 genes of herpes simplex virus.";
RL VIROLOGY 210:100-108(1995).
RN
RN
RP SEQUENCE FROM N.A.
RC STRAIN=JURA;
RA SCHWYZER M.;
RA LOWERY D.E., SIMARD C., BELLO L.J., THIRY E., VLCEK C.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN
RN
RP SEQUENCE FROM N.A.
RC STRAIN=JURA;
RA SCHWYZER M.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; 248053; G971327;
DR EMBL; AJ004801; E1187337;
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 299 AA; 32379 MW; A3897D4A CRC32;

Query Match 16.7%; Score 92; DB 14; Length 299;
Best Local Similarity 47.2%; Pred. No. 8.15e-02;
Matches 17; Conservative 9; Mismatches 6; Indels 4; Gaps 4;

Db 45 PRFVCEVEIPAGPTFTSSSITHLRV-EPSTGALL 79
QY 41 PSYV-KFQDNPVGSQTF-SAGL-HLRVDPSTGALV 73

RESULT 4
ID Q65581 PRELIMINARY; PRT; 300 AA.
AC Q65581;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
```

```
DE UL7 POLYPEPTIDE.
GN UL7.
OS BOVINE HERPESVIRUS TYPE 1.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.
RN
RP SEQUENCE FROM N.A.
RC STRAIN=SCHOENBOEKEN;
RX MEDLINE; 96135223.
RA SCHWITT J., KEIL G.M.;
RT "Identification and characterization of the bovine herpesvirus 1 UL7
RT gene and gene product which are not essential for virus replication
RT in cell culture.";
RL J. VIROL. 70:1091-1099(1996).
DR EMBL; X91751; G1006630;
SQ SEQUENCE 300 AA; 32450 MW; 86377347 CRC32;

Query Match 16.7%; Score 92; DB 14; Length 300;
Best Local Similarity 47.2%; Pred. No. 8.15e-02;
Matches 17; Conservative 9; Mismatches 6; Indels 4; Gaps 4;

Db 46 PRFVCEVEIPAGPTFTSSSITHLRV-EPSTGALL 80
QY 41 PSYV-KFQDNPVGSQTF-SAGL-HLRVDPSTGALV 73

RESULT 5
ID O84354 PRELIMINARY; PRT; 566 AA.
AC O84354;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 63.5 KD PROTEIN.
GN Cr350.
OS CHLAMYDIA TRACHOMATIS.
OC BACTERIA; CHLAMYDIALES; CHLAMYDIACEAE; CHLAMYDIA.
RN
RN
RP SEQUENCE FROM N.A.
RC STRAIN=D/UV-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATSOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RT "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
RT Chlamydia trachomatis.";
RL SCIENCE 0:0-0(1998).
RN
RN
RP SEQUENCE FROM N.A.
RC STRAIN=D/UV-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
RA MITCHELL W.P., OLINGER L., TATSOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AE001308; G3328771;
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 566 AA; 63508 MW; C1A1C491 CRC32;

Query Match 16.2%; Score 89; DB 2; Length 566;
Best Local Similarity 26.2%; Pred. No. 2.28e-01;
Matches 17; Conservative 25; Mismatches 18; Indels 5; Gaps 5;

Db 63 IAESYLOQSFSEPTYIR-KSAITGA-GLSGSSEA-LELLSEAIETQDLYEQL-L-ILNA 117
QY 8 FGCGYVQTPFLSENSVRYKISAGSCPLSTAGPSYKFDQNPVGSQTFSSAGLHLRVDP 67
Db 118 ATSL 122
QY 68 STGAL 72

RESULT 6
ID Q81862 PRELIMINARY; PRT; 1693 AA.
AC Q81862;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
```

```

01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
ORF 1.
OS HEPATITIS E VIRUS (HEV).
OC VIRUSES: SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALCIVIRIDAE;
OC CALCIVIRUS.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE; 92024067.
RA TAM A.W., SMITH M.M., GUERRA M.E., HUANG C.C., BRADLEY D.W., FRY K.E.,
RA REYES G.R.;
RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT full-length viral genome.";
RL VIROLOGY 185:120-131(1991).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE; 92261377.
RA UCHIDA T., SUZUKI K., HAYASHI N., IIDA F., HARA T., OO S.S.,
RA WANG C.K., SHIRATA T., ICHIKAWA M., RIKIHISA T., MIZUNO K., WIN K.M.,
RT "Hepatitis E virus: cdna cloning and expression.";
RL MICROBIOL. IMMUNOL. 36:67-79(1992).
RN [3]
RN SEQUENCE FROM N.A.
RX MEDLINE; 92335008.
RA AYE T.T., UCHIDA T., MA X.Z., IIDA F., SHIRATA T., ZHUANG H.,
RA WIN K.M.;
RT "Complete nucleotide sequence of a hepatitis E virus isolated from
RT the Xinjiang epidemic (1986-1988) of China.";
RL NUCLEIC ACIDS RES. 20:3512-3512(1992).
RN [4]
RN SEQUENCE FROM N.A.
RX MEDLINE; 92115700.
RA TSAREV S.A., EMERSON S.U., REYES G.R., TSAREVA T.S., LEGTERS L.J.,
RA MALIK I.A., IQBAL M., PURCELL R.H.;
RT "Characterization of a prototype strain of hepatitis E virus.";
RL PROC. NATL. ACAD. SCI. U.S.A. 89:559-563(1992).
RN [5]
RN SEQUENCE FROM N.A.
RX MEDLINE; 92271462.
RA FRY K.E., TAM A.W., SMITH M.M., KIM J.P., LUK K.C., YOUNG L.M.,
RA PIATAK M., FELDMAN R.A., YUN K.Y., PURDY M.A., ET AL.;
RT "Hepatitis E virus (HEV): strain variation in the nonstructural gene
RT region encoding consensus motifs for an RNA-dependent RNA polymerase
RT and an ATP/GTP binding site.";
RL VIRUS GENES 6:173-185(1992).
RN [6]
RN SEQUENCE FROM N.A.
RX MEDLINE; 93348763.
RA BI S.L., PURDY M.A., MCCAUSTLAND K.A., MARGOLIS H.S., BRADLEY D.W.;
RT "The sequence of hepatitis E virus isolated directly from a single
RT source during an outbreak in China.";
RL VIRUS RES. 28:233-247(1993).
DR ENBL: L08816: G33009;
SQ SEQUENCE 1693 AA: 185052 MW: 30A917E1 CRC32;

Query Match 15.6%; Score 86; DB 14; Length 1693;
Best Local Similarity 25.4%; Pred. No. 6.27e-01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps

Db 572 FRITSFVDCAVLENGPERINLSEFDSQSQTMAAGPSLTYAASAGLEVVRYVAAGLDHRV 631
QY 8 FGQGYVQTPFLSESNVRYKIAGSCPLSTAGSPSYVVFQDNPVGQSOT-F-SAGLHLR-V 64
Db 632 FAPGVSP 638
QY 65 FDPSTGA 71

RESULT 7
ID Q81344 PRELIMINARY; PRT: 1693 AA.
AC Q81344;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

```



```
OC VIRUSES: DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; SIPHOVIRIDAE;
OC LAMBDA PHAGE GROUP.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94374705.
RA HARTLEY N.M., MURPHY G.O., BRUTON C.J., CHATER K.F.:
RT "Sequence of the essential early region of phi C31, a temperate phage
RT of Streptomyces spp. with unusual features in its lytic
RT development.";
RL GENE 147:29-40(1994).
DR EMBL: X75288; G579071; -.
SQ SEQUENCE 281 AA; 31680 MW; 0E734629 CRC32;

Query Match 15.1%; Score 83; DB 9; Length 281;
Best Local Similarity 24.1%; Pred. No. 1.68e+00;
Matches 14; Conservative 17; Mismatches 24; Indels 3; Gaps 3;

Db 194 AYMNADFIIDPGNREPMPEFDGAALVHTDETWA-EKPVETGPDVFAQLHLRQTFD 250
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 11 GYVOTPELSNSVRYKIS-IAGSCPLSTAGPSYVKFQDNPVGSQTSAGLHLR-VFD 66

RESULT 14
ID Q54178 PRELIMINARY; PRT; 1577 AA.
AC Q54178; Q54247.
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE DE GLUCOSYLTRANSFERASE.
GN GTFG.
OS STREPTOCOCCUS GORDONII.
OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CHALLIS;
RX MEDLINE: 96157084.
RA VICKERMAN M.M., SULAVIK M.C., CLEWELL D.B.:
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase
RT phase variants.";
RL DEV. BIOL. STAND. 85:309-314(1995).
RN [2]
RP SEQUENCE OF 1-96 FROM N.A.
RX STRAIN=CHALLIS;
RX MEDLINE: 92276337.
RA SULAVIK M.C., TARDIF G., CLEWELL D.B.:
RT "Identification of a gene, rgg, which regulates expression of
RT glucosyltransferase and influences the spp phenotype of Streptococcus
RT gordonii Challis.";
RL J. BACTERIOL. 174:3577-3586(1992).
DR EMBL: U12643; G1054877; -.
DR EMBL: M89776; G153795; -.
DR PFAM: PF00128; alpha-amylase; 1.
KW TRANSFERASE.
SQ SEQUENCE 1577 AA; 177805 MW; 7D58FEC2 CRC32;

Query Match 15.1%; Score 83; DB 2; Length 1577;
Best Local Similarity 41.2%; Pred. No. 1.68e+00;
Matches 14; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Db 239 QDDGTVKKNFAVELNGKILYFDAETGALVDSNEY 272
   : : : : : : : : : : : : : : : : : : : : : :
QY 47 QDNPVGSQTSAGLHLRV--FDPSTGALVDSKSY 78

RESULT 15
ID O39221 PRELIMINARY; PRT; 1693 AA.
AC O39221.
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE NONSTRUCTURAL POLYPROTEIN.
OS HEPATITIS E VIRUS (HEV).
```

```
OC VIRUSES: SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; CALCIVIRIDAE;
OC CALCIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-HYDERABAD, INDIA;
RA ANSARI I.H., NANDA S.K., DURGAPAL H., JAMEEL S., PANDA S.K.:
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA ANSARI I.H., NANDA S.K., DURGAPAL H., JAMEEL S., PANDA S.K.:
RT "Translational analysis of complete HEV genome.";
RL SUBMITTED (JUL-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF028091; G2565434; -.
DR EMBL: AF076239; G3328379; -.
KW POLYPROTEIN.
SQ SEQUENCE 1693 AA; 185620 MW; E8D755D2 CRC32;

Query Match 15.1%; Score 83; DB 14; Length 1693;
Best Local Similarity 23.9%; Pred. No. 1.68e+00;
Matches 16; Conservative 23; Mismatches 25; Indels 3; Gaps 3;

Db 572 FRTSFVDGAVLETPGPERHNLSDASOSTMAAGPFLTYAASAAGLEVRYVYGAGLDHRAI 631
   : : : : : : : : : : : : : : : : : : : : : :
QY 8 FGQGYVOTPELSNSVRYKISITAGSCPLSTAGPSYVKFQDNPVGSQT-F-SAGLHLR-V 64

Db 632 FAPGVSP 638
   : : :
QY 65 FDPSTGA 71
```

Search completed: Wed Sep 1 16:13:32 1999
Job time : 27 secs.

WATERMAN

***** (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:29:24 1999; MasPar time 5.26 Seconds
Tabular output not generated. 145.242 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (66-75) from PCTUS9913024.pep (11 of 12)
Perfect Score: 86
Sequence: 1 DPSTGALVDSKSYA 14

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spiremb19

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 24.488; Variance 26.991; scale 0.907

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	59	68.6	1577	2	GLUCOSYLTRANSFERASE	7.85e-02
2	56	65.1	368	2	HEAT SHOCK PROTEIN/SE	4.07e-01
3	56	65.1	472	2	SERINE PROTEASE.	4.07e-01
4	52	60.5	405	4	KIAA0574 PROTEIN (FRAG	3.35e+00
5	52	60.5	552	5	RADIAL SPOKEHEAD.	3.35e+00
6	52	60.5	870	4	KIAA0563 PROTEIN.	3.35e+00
7	51	59.3	540	5	PHOTOLYASE.	5.58e+00
8	51	59.3	1340	14	POLYPROTEIN (FRAGMENT)	5.58e+00
9	51	59.3	3493	14	POLYPROTEIN.	5.58e+00
10	50	58.1	142	14	ENVELOPE GLYCOPROTEIN	9.22e+00
11	50	58.1	156	11	MONOCARBOXYLATE TRANSP	9.22e+00
12	50	58.1	262	2	W-AMINO-TRANSFERASE-LI	9.22e+00
13	50	58.1	371	2	CMP-N-ACETYLNEURAMINAT	9.22e+00
14	50	58.1	492	11	MONOCARBOXYLATE TRANSP	9.22e+00
15	50	58.1	542	13	RETINAL EPITHELIAL MEM	9.22e+00
16	50	58.1	542	13	SIMILARITY TO MYOSIN H	9.22e+00
17	50	58.1	1133	10	PATERNALLY EXPRESSED P	9.22e+00
18	50	58.1	1378	11	YCDH.	1.51e+01
19	49	57.0	319	2	HEMAGGLUTININ (FRAGMEN	1.51e+01
20	49	57.0	327	14	HEMAGGLUTININ (FRAGMEN	1.51e+01

21	49	57.0	329	14	HEMAGGLUTININ (FRAGMEN	1.51e+01
22	49	57.0	329	14	HEMAGGLUTININ (FRAGMEN	1.51e+01
23	49	57.0	427	5	PARASITOPHOROUS VACUOL	1.51e+01
24	49	57.0	734	11	INSULIN RECEPTOR SUBST	1.51e+01
25	49	57.0	966	1	HYPOTHETICAL 104.9 KD	1.51e+01
26	49	57.0	984	5	SERINE RICH PROTEIN (S	1.51e+01
27	49	57.0	1139	2	ANCHOR PROTEIN, LCM.	1.51e+01
28	48	55.8	195	2	HYPOTHETICAL 20.8 KD P	2.46e+01
29	48	55.8	328	14	HAEMAGGLUTININ (HAI DO	2.46e+01
30	48	55.8	329	14	HEMAGGLUTININ (FRAGMEN	2.46e+01
31	48	55.8	329	14	HEMAGGLUTININ (FRAGMEN	2.46e+01
32	48	55.8	329	14	HEMAGGLUTININ (FRAGMEN	2.46e+01
33	48	55.8	329	14	HEMAGGLUTININ (FRAGMEN	2.46e+01
34	48	55.8	329	14	HEMAGGLUTININ (FRAGMEN	2.46e+01
35	48	55.8	329	14	HEMAGGLUTININ (FRAGMEN	2.46e+01
36	48	55.8	427	10	S-LOCUS-SPECIFIC GLYCO	2.46e+01
37	48	55.8	477	11	SCK, PARTIAL CDS (FRAG	2.46e+01
38	48	55.8	480	2	PUTATIVE TRANSCRIPTION	2.46e+01
39	48	55.8	540	4	SCK, PARTIAL CDS (FRAG	2.46e+01
40	48	55.8	550	14	HAEMAGGLUTININ PRECURS	2.46e+01
41	48	55.8	580	3	DIHYDROXYACETONE KINAS	2.46e+01
42	48	55.8	712	2	OMP1D PRECURSOR.	2.46e+01
43	48	55.8	1571	11	ZINC FINGER PROTEIN.	2.46e+01
44	47	54.7	329	14	HEMAGGLUTININ (FRAGMEN	3.97e+01
45	47	54.7	329	14	HEMAGGLUTININ (FRAGMEN	3.97e+01

ALIGNMENTS

RESULT	1
ID	Q54178 PRELIMINARY; PRT: 1577 AA.
AC	Q54178; Q54247;
DT	01-NOV-1996 (TREMBREL. 01, CREATED)
DT	01-NOV-1996 (TREMBREL. 01, LAST SEQUENCE UPDATE)
DT	01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE	GLUCOSYLTRANSFERASE.
GN	GTFG.
OS	STREPTOCOCCUS GORDONII.
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC	STREPTOCOCCUS.
RN	[1]
RC	SEQUENCE FROM N.A.
RP	SEQUENCE OF 1-96 FROM N.A.
RX	STRAIN-CHALLIS;
RX	MEDLINE: 92276337
RA	SULAVIK M.C., TARDIF G., CLEWELL D.B.;
RT	*Identification of a gene, rgg, which regulates expression of
RT	glucosyltransferase and influences the Spp phenotype of Streptococcus
RT	gordonii Challis.*;
RL	J. BACTERIOL. 174:3577-3586(1992).
DR	EMBL: U12643; G1054877; -
DR	EMBL: M89776; G153795; -
DR	PFAM: PF00128; alpha-amylase; 1.
KW	TRANSFERASE.
SQ	SEQUENCE 1577 AA; 177805 MW; 7D58FEC2 CRC32;

Query Match 58.6%; Score 59; DB 2; Length 1577;
Best Local Similarity 69.2%; Pred. No. 7.85e-02;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 260 DAETGALVDSNEY 272

QY 66 DPSTGALVDSKSY 78

RESULT 2

```

ID Q46094      PRELIMINARY;      PRT;      368 AA.
AC Q46094;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HEAT SHOCK PROTEIN/SERINE PROTEASE (FRAGMENT).
GN HTA.
OS CAMPYLOBACTER JEJUNI.
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; CAMPYLOBACTER GROUP;
OC CAMPYLOBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UA580;
RX MEDLINE; 90384493.
RA TAYLOR D.E., HIRATSUKA K.;
RT "Use of non-radioactive DNA probes for detection of Campylobacter
RT jejuni and Campylobacter coli in stool specimens.";
RL MOL. CELL. PROBES 4:261-271(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-UA580;
RA HIRATSUKA K.;
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U27271; G881375;
DR PFAM; PF00089; trypsin; 1.
DR PFAM; PF00595; PDZ; 2.
KW HEAT SHOCK; PROTEASE.
FT NON_TER
FT NON_TER
SQ SEQUENCE 368 AA; 39491 MW; A5062589 CRC32;

Query Match      65.1%; Score 56; DB 2; Length 368;
Best Local Similarity 72.7%; Pred. No. 4.07e-01;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 121 SGGALVDSRGY 131
QY 68 STGALVDSKSY 78

RESULT 3
ID Q46120      PRELIMINARY;      PRT;      472 AA.
AC Q46120;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SERINE PROTEASE.
GN HTA.
OS CAMPYLOBACTER JEJUNI.
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; CAMPYLOBACTER GROUP;
OC CAMPYLOBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-81116;
RA HENDERSON J., WOOD A.C., EMERY M.J., WREN B.W., KETLEY J.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X82628; E315368;
DR PFAM; PF00089; trypsin; 1.
DR PFAM; PF00595; PDZ; 2.
KW PROTEASE.
SQ SEQUENCE 472 AA; 50940 MW; 6BCA101 CRC32;

Query Match      65.1%; Score 56; DB 2; Length 472;
Best Local Similarity 72.7%; Pred. No. 4.07e-01;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 225 SGGALVDSRGY 235
QY 68 STGALVDSKSY 78

RESULT 4
ID O60320      PRELIMINARY;      PRT;      405 AA.
AC O60320;

```

```

DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE KIAA0574 PROTEIN (FRAGMENT).
GN KIAA0574.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 98290545.
RA NAGASE T., ISHIKAWA K., MIYAJIMA N., TANAKA A., KOTANI H., NOMURA N.,
RA OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA RES. 5:31-39(1998).
DR EMBL; AB011146; D1026430;
FT NON_TER
FT NON_TER
SQ SEQUENCE 405 AA; 42409 MW; 09BA9932 CRC32;

Query Match      60.5%; Score 52; DB 4; Length 405;
Best Local Similarity 77.8%; Pred. No. 3.35e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 345 SLVDSKAYA 353
QY 71 ALVDSKSYA 79

RESULT 5
ID O46178      PRELIMINARY;      PRT;      552 AA.
AC O46178;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE RADIAL SPOKEHEAD.
OS STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; EUECHINOIDEA;
OC ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE; STRONGYLOCENTROTUS.
RN [1]
RP SEQUENCE FROM N.A.
RX GINGRAS D., GAGNON C.;
RT "Molecular cloning and characterization of a radial spoke head
RT protein of sea urchin sperm axonemes: involvement of the protein in
RT the regulation of sperm motility.";
RL MOL. BIOL. CELL 9:513-522(1998).
DR EMBL; U73123; G2905895;
SQ SEQUENCE 552 AA; 62723 MW; 898CFCCC CRC32;

Query Match      60.5%; Score 52; DB 5; Length 552;
Best Local Similarity 50.0%; Pred. No. 3.35e+00;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 39 PTEALVNKAY 50
QY 67 PSTGALVDSKSY 78

RESULT 6
ID O60309      PRELIMINARY;      PRT;      870 AA.
AC O60309;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE KIAA0563 PROTEIN.
GN KIAA0563.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 98290545.
RA NAGASE T., ISHIKAWA K., MIYAJIMA N., TANAKA A., KOTANI H., NOMURA N.,
RA OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT code for large proteins in vitro.";
RL DNA RES. 5:31-39(1998).
DR EMBL; AB011135; D1028419; -.
SQ SEQUENCE 870 AA; 94665 MW; EB9CD3F4 CRC32;

Query Match
Best Local Similarity 60.5%; Score 52; DB 4; Length 870;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 798 EPAQDSLVSESYT 811
QY 66 DPSTGALVDSKSYA 79
      :|: |||:|:|:
      || |||:|:|:

RESULT 7
ID Q24281 PRELIMINARY; PRT; 540 AA.
AC Q24281;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PHOTOLYASE.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PHR-; TISSUE-OVARY;
RX MEDLINE; 96178677.
RA TODO T., RYO H., YAMAMOTO K., TOH H., INUI T., AYAKI H., NOMURA T.,
RA IKENAGA M.;
RT "Similarity among the Drosophila (6-4)photolyase, a human photolyase
RT homolog, and the DNA photolyase-blue-light photoreceptor family.";
RL SCIENCE 272:109-112(1996).
DR EMBL; D83701; D1012739; -.
KW LYASE.
SQ SEQUENCE 540 AA; 62548 MW; 62F139EE CRC32;

Query Match
Best Local Similarity 59.3%; Score 51; DB 5; Length 540;
Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 456 EPWKASLVDRAYG 469
QY 66 DPSTGALVDSKSYA 79
      :|: |||:|:|:
      || |||:|:|:

RESULT 8
ID Q88274 PRELIMINARY; PRT; 1340 AA.
AC Q88274;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE POLYPROTEIN (FRAGMENT).
OS SWEET POTATO FEATHERY MOTTLE VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
OC POTYVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SEVERE;
RX MEDLINE; 95390803.
RA MORI M., SAKAI J., KIMURA T., USUGI T., HAYASHI T., HANADA K.,
RA NISHIGUCHI M.;
RT "Nucleotide sequence analysis of two nuclear inclusion body and coat
RT protein genes of a sweet potato feathery mottle virus severe strain
RT (SPFMV-S) genomic RNA.";

```

```

RL ARCH. VIROL. 140:1473-1482(1995).
DR EMBL; D38543; D1008127; -.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00767; Poty_coat; 1.
DR PFAM; PF00863; Peptidase_C4; 1.
KW POLYPROTEIN; COAT PROTEIN.
FT NON_TER 1
FT CHAIN 1 16 CYTOPLASMIC INCLUSION PROTEIN.
FT CHAIN 17 69 6K2 PROTEIN.
FT CHAIN 70 261 VPG PROTEIN.
FT CHAIN 262 504 NTA PROTEASE.
FT CHAIN 505 1025 NUCLEAR INCLUSION PROTEIN B.
FT CHAIN 1026 1340 COAT PROTEIN.
SQ SEQUENCE 1340 AA; 151873 MW; 12C01867 CRC32;

Query Match
Best Local Similarity 59.3%; Score 51; DB 14; Length 1340;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 146 DPLTGAVIDSPYT 159
QY 66 DPSTGALVDSKSYA 79
      || |||:|:|:
      || |||:|:|:

RESULT 9
ID O39734 PRELIMINARY; PRT; 3493 AA.
AC O39734;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POLYPROTEIN.
OS SWEET POTATO FEATHERY MOTTLE VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; POTYVIRIDAE;
OC POTYVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S STRAIN;
RA NISHIGUCHI M.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S STRAIN;
RX MEDLINE; 98336489.
RA SAKAI J., MORI M., MORISHITA T., TANAKA M., HANADA K., USUGI T.,
RA NISHIGUCHI M.;
RT "Complete nucleotide sequence and genome organization of sweet potato
RT feathery mottle virus (S strain) genomic RNA: the large coding region
RT of the P1 gene.";
RL ARCH. VIROL. 142:1553-1562(1997).
DR EMBL; D86371; D1023570; -.
DR PFAM; PF00271; helicase_C; 1.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00767; Poty_coat; 1.
DR PFAM; PF00851; Peptidase_C6; 1.
DR PFAM; PF00863; Peptidase_C4; 1.
KW POLYPROTEIN.
FT CHAIN 1 664 POTENTIAL.
FT CHAIN 665 1122 POTENTIAL.
FT CHAIN 1123 1474 POTENTIAL.
FT CHAIN 1475 1526 POTENTIAL.
FT CHAIN 1527 2169 POTENTIAL.
FT CHAIN 2170 2222 POTENTIAL.
FT CHAIN 2223 2414 GENOME-LINKED VIRAL PROTEIN.
FT CHAIN 2415 2657 PROTEASE.
FT CHAIN 2658 3178 REPLICASE.
FT CHAIN 3179 3493 COAT PROTEIN.
SQ SEQUENCE 3493 AA; 393818 MW; A02EFA05 CRC32;

Query Match
Best Local Similarity 59.3%; Score 51; DB 14; Length 3493;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 2299 DPLTGAVIDSPYT 2312

```

QY	66 DPSTGALVDSKSYA 79
RESULT 10	PRELIMINARY; PRT; 142 AA.
ID Q87600	
AC Q87600	SEQUENCE FROM N.A.
DT 01-NOV-1996	(TREMBLEL. 01, CREATED)
DT 01-NOV-1996	(TREMBLEL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998	(TREMBLEL. 08, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).	
GN ENV.	
OS SIMIAN IMMUNODEFICIENCY VIRUS (SIV-AGM).	
OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; LENTIVIRUS.	
RN [1]	
RC STRAIN-P081	
RA BIBOLLET-RUCHE F., BRENGUES C., GALAT-LUONG A., GALAT G., POURRUT X., VIDAL N., VEAS F., DURAND J.P., CUNY G.;	
RL J. VIROL. 71:307-313(1997).	
DR EMBL; U37209; G1353456; -	
DR PFAM; PF00517; GP41; 1.	
KW ENVELOPE PROTEIN.	
FT NON_TER 1	
FT NON_TER 142	
FT NON_TER 142	
SQ SEQUENCE 142 AA; 16852 MW; C58A9A91 CRC32;	
Query Match	58.1%; Score 50; DB 14; Length 142;
Best Local Similarity	46.7%; Pred. No. 9.22e+00;
Matches	7; Conservative 5; Mismatches 2; Indels 1; Gaps 1;
Db 86 EPDGGSLRSRYSVT 100	
: : : :	
QY 66 DPSTGALV-DSKSYA 79	
RESULT 11	PRELIMINARY; PRT; 156 AA.
ID Q35308	
AC Q35308	SEQUENCE FROM N.A.
DT 01-JAN-1998	(TREMBLEL. 05, CREATED)
DT 01-JAN-1998	(TREMBLEL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998	(TREMBLEL. 05, LAST ANNOTATION UPDATE)
DE MONOCARBOXYLATE TRANSPORTER MCT3 (FRAGMENT).	
OS MUS MUSCULUS (MOUSE).	
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;	
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.	
RN [1]	
RC STRAIN-C3H;	
RA PHILP N.O., YOON H.;	
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.	
DR EMBL; AF019111; G2407664; -	
FT NON_TER 1	
FT NON_TER 1	
SQ SEQUENCE 156 AA; 16253 MW; DB394B9E CRC32;	
Query Match	58.1%; Score 50; DB 11; Length 156;
Best Local Similarity	61.5%; Pred. No. 9.22e+00;
Matches	8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;
Db 37 PSAGRLVDALKNY 49	
: : : :	
QY 67 PSTGALVDS-KSY 78	
RESULT 12	PRELIMINARY; PRT; 262 AA.
ID Q44188	
AC Q44188	SEQUENCE FROM N.A.
DT 01-NOV-1996	(TREMBLEL. 01, CREATED)
DT 01-NOV-1996	(TREMBLEL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998	(TREMBLEL. 08, LAST ANNOTATION UPDATE)
DE W-AMINO-TRANSFERASE-LIKE PROTEIN.	
GN OATA.	
OS AGROBACTERIUM RADIOBACTER.	

QY	66 DPSTGALVDSKSYA 79	11 11111111 1111
RESULT 10	PRELIMINARY; PRT; 142 AA.	
ID 087600	SEQUENCE FROM N.A.	
AC 087600	SEQUENCE FROM N.A.	
DT 01-NOV-1996	(TREMUREL. 01, CREATED)	
DT 01-NOV-1996	(TREMUREL. 01, LAST SEQUENCE UPDATE)	
DT 01-NOV-1998	(TREMUREL. 08, LAST ANNOTATION UPDATE)	
DE	ENVELOPE GLYCOPROTEIN (FRAGMENT).	
GN ENV		
OS	SIMIAN IMMUNODEFICIENCY VIRUS (SIV-AGM).	
OC	VIRUSES; RETROVIRUSES; RETROVIRIDAE; LENTIVIRUS.	
RC	SEQUENCE FROM N.A.	
RP	STRAIN-P081	
RA	BIBOLLET-RUCHE F., BRENGUES C., GALAT-LUONG A., GALAT G., POURRUT X., VIDAL N., VEAS F., DURAND J.P., CUNY G.	
RL	J. VIROL. 71:307-313(1997).	
DR	EMBL; U37209; G1353456; -	
DR	PFAM; PF00517; GP41; 1.	
KW	ENVELOPE PROTEIN.	
FT	NON_TER 1	
FT	NON_TER 142	
FT	NON_TER 142	
SEQ	SEQUENCE 142 AA; 16852 MW; C58A9A91 CRC32;	
Query Match	58.1%; Score 50; DB 14; Length 142;	
Best Local Similarity	46.7%; Pred. No. 9.22e+00;	
Matches	7; Conservative 5; Mismatches 2; Indels 1; Gaps 1;	
Db	86 EPDSGGLSRDSRST 100	
QY	66 DPSTGALV-DSKSYA 79	
11 11111111 1111		
RESULT 11	PRELIMINARY; PRT; 156 AA.	
ID 035308	SEQUENCE FROM N.A.	
AC 035308	SEQUENCE FROM N.A.	
DT 01-JAN-1998	(TREMUREL. 05, CREATED)	
DT 01-JAN-1998	(TREMUREL. 05, LAST SEQUENCE UPDATE)	
DT 01-JAN-1998	(TREMUREL. 05, LAST ANNOTATION UPDATE)	
DE	MONOCARBOXYLATE TRANSPORTER MCT3 (FRAGMENT).	
OS	MUS MUSCULUS (MOUSE).	
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;	
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.	
RC	SEQUENCE FROM N.A.	
RP	STRAIN-C3H;	
RA	PHILP N.O., YOON H.	
RL	SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.	
DR	EMBL; AF019111; G2407664; -	
FT	NON_TER 1	
FT	NON_TER 1	
SEQ	SEQUENCE 156 AA; 16253 MW; DB394B9E CRC32;	
Query Match	58.1%; Score 50; DB 11; Length 156;	
Best Local Similarity	61.5%; Pred. No. 9.22e+00;	
Matches	8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;	
Db	37 PSAGRLVDALKNY 49	
QY	67 PSTGALVDS-KSY 78	
11 11111111 1111		
RESULT 12	PRELIMINARY; PRT; 262 AA.	
ID 044188	SEQUENCE FROM N.A.	
AC 044188	SEQUENCE FROM N.A.	
DT 01-NOV-1996	(TREMUREL. 01, CREATED)	
DT 01-NOV-1996	(TREMUREL. 01, LAST SEQUENCE UPDATE)	
DT 01-NOV-1998	(TREMUREL. 08, LAST ANNOTATION UPDATE)	
DE	W-AMINO-TRANSFERASE-LIKE PROTEIN.	
GN OATA		
OS	AGROBACTERIUM RADIOBACTER.	

QY	66 DPSTGALVDSKSYA 79
RESULT 10	PRELIMINARY; PRT; 142 AA.
ID Q87600	
AC Q87600	SEQUENCE FROM N.A.
DT 01-NOV-1996	(TREMBLEL. 01, CREATED)
DT 01-NOV-1996	(TREMBLEL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998	(TREMBLEL. 08, LAST ANNOTATION UPDATE)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).	
GN ENV.	
OS SIMIAN IMMUNODEFICIENCY VIRUS (SIV-AGM).	
OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; LENTIVIRUS.	
RN [1]	
RC STRAIN-P081	
RA BIBOLLET-RUCHE F., BRENGUES C., GALAT-LUONG A., GALAT G., POURRUT X., VIDAL N., VEAS F., DURAND J.P., CUNY G.;	
RL J. VIROL. 71:307-313(1997).	
DR EMBL; U37209; G1353456; -	
DR PFAM; PF00517; GP41; 1.	
KW ENVELOPE PROTEIN.	
FT NON_TER 1	
FT NON_TER 142	
FT NON_TER 142	
SQ SEQUENCE 142 AA; 16852 MW; C58A9A91 CRC32;	
Query Match	58.1%; Score 50; DB 14; Length 142;
Best Local Similarity	46.7%; Pred. No. 9.22e+00;
Matches	7; Conservative 5; Mismatches 2; Indels 1; Gaps 1;
Db 86 EPDGGSLRSRVSYSVT 100	
: : : : : : : : : : : : :	
QY 66 DPSTGALV-DSKSYA 79	
RESULT 11	PRELIMINARY; PRT; 156 AA.
ID O35308	
AC O35308	SEQUENCE FROM N.A.
DT 01-JAN-1998	(TREMBLEL. 05, CREATED)
DT 01-JAN-1998	(TREMBLEL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998	(TREMBLEL. 05, LAST ANNOTATION UPDATE)
DE MONOCARBOXYLATE TRANSPORTER MCT3 (FRAGMENT).	
OS MUS MUSCULUS (MOUSE).	
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;	
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.	
RN [1]	
RC STRAIN-C3H;	
RA PHILP N.O., YOON H.;	
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.	
DR EMBL; AF019111; G2407664; -	
FT NON_TER 1	
FT NON_TER 1	
SQ SEQUENCE 156 AA; 16253 MW; DB394B9E CRC32;	
Query Match	58.1%; Score 50; DB 11; Length 156;
Best Local Similarity	61.5%; Pred. No. 9.22e+00;
Matches	8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;
Db 37 PSAGRLVDALKNY 49	
: : : : : : : : :	
QY 67 PSTGALVDS-KSY 78	
RESULT 12	PRELIMINARY; PRT; 262 AA.
ID O44188	
AC O44188	SEQUENCE FROM N.A.
DT 01-NOV-1996	(TREMBLEL. 01, CREATED)
DT 01-NOV-1996	(TREMBLEL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998	(TREMBLEL. 08, LAST ANNOTATION UPDATE)
DE W-AMINO-TRANSFERASE-LIKE PROTEIN.	
GN OATA.	
OS AGROBACTERIUM RADIOBACTER.	

Query Match 58.1%; Score 50; DB 2; Length 371;
Best Local Similarity 53.8%; Pred. No. 9.22e+00;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 164 DGGTGNLIQSSY 176
QY 66 DPSTGALVDSKY 78
I:||||:|I|

RESULT 14
ID 070461 PRELIMINARY; PRT; 492 AA.
AC 070461;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE MONOCARBOXYLATE TRANSPORTER MCT3.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-RETINA;
RA PHILP N.J., YOON H.;
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR ENBL; AF059258; G3170609; -.
SQ SEQUENCE 492 AA; 51590 MW; 20F051C1 CRC32;

Query Match 58.1%; Score 50; DB 11; Length 492;
Best Local Similarity 61.5%; Pred. No. 9.22e+00;
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 373 PSAGRLVDALKNY 385
QY 67 PSTGALVDS-KSY 78
I:||||:|I|

RESULT 15
ID 013151 PRELIMINARY; PRT; 542 AA.
AC 013151;
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE MONOCARBOXYLATE TRANSPORTER 3.
GN MCT3.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LEGHORN;
RA PHILP N.J., YOON H.;
RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR ENBL; AF000240; G2198807; -.
SQ SEQUENCE 542 AA; 58085 MW; 836DA7B1 CRC32;

Query Match 58.1%; Score 50; DB 13; Length 542;
Best Local Similarity 61.5%; Pred. No. 9.22e+00;
Matches 8; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

Db 410 PSAGRLVDALKNY 422
QY 67 PSTGALVDS-KSY 78
I:||||:|I|

Search completed: Wed Sep 1 16:29:45 1999
Job time : 21 secs.

(TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:33:07 1999; MasPar time 4.43 Seconds
62.340 Million cell updates/sec

Tabular output not generated.

Title: >PCT-US99-13024-2

Description: (81-93) from PCTUS9913024.pep (12 of 12)

Perfect Score: 68
Sequence: 1 STSNDTSAAFVS 13

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 15.573; Variance 46.313; scale 0.336

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	54	79.4	W63157	Human calcium channel	1.78e+01
2	54	79.4	W72605	Human calcium channel	1.78e+01
3	54	79.4	W72606	Human calcium channel	1.78e+01
4	54	79.4	W63158	Human calcium channel	1.78e+01
5	54	79.4	R39697	Myasthenic antigenic	1.78e+01
6	54	79.4	R39563	Human neuronal VDCC b	1.78e+01
7	54	79.4	R39563	Human calcium channel	1.78e+01
8	54	79.4	R72611	Human calcium channel	1.78e+01
9	54	79.4	R39551	Sequence of splice va	1.78e+01
10	54	79.4	R72604	Human calcium channel	1.78e+01
11	54	79.4	W63156	Human calcium channel	1.78e+01
12	54	79.4	W63147	Human calcium channel	1.78e+01
13	54	79.4	R39565	Human neuronal VDCC b	1.78e+01
14	54	79.4	W63144	Human calcium channel	1.78e+01
15	54	79.4	W63144	Human calcium channel	1.78e+01
16	54	79.4	R39552	Sequence of splice va	1.78e+01

17	54	79.4	660 14	R72613	Human neuronal calciu	1.78e+01
18	44	64.7	636 36	W79589	Human kv potassium ch	2.29e+02
19	44	64.7	636 36	W79591	Human kv potassium ch	2.29e+02
20	44	64.7	655 36	W79590	Human kv potassium ch	2.29e+02
21	44	64.7	1130 15	R71729	Merosin major subunit	2.29e+02
22	44	64.7	1130 3	R13436	Merosin M polypeptide	2.29e+02
23	44	64.7	3110 15	R71730	Merosin major subunit	2.29e+02
24	43	63.2	757 20	W03179	Bovine poly-immunoglo	2.93e+02
25	42	61.8	127 28	W31716	Gamma-1 heavy chain a	3.74e+02
26	42	61.8	653 21	W14264	Z. japonica phosphoen	3.74e+02
27	42	61.8	953 19	R97284	Human 26S proteasome	3.74e+02
28	42	61.8	2783 5	R23963	APP-1 (Ala 2460 Val)	3.74e+02
29	42	61.8	2783 5	R23962	APP-1	3.74e+02
30	41	60.3	98 13	R72068	Dp10 VH region.	4.76e+02
31	41	60.3	120 25	W27550	Human Ab heavy chain	4.76e+02
32	41	60.3	120 10	R54795	SPA-reactive IgM heav	4.76e+02
33	41	60.3	123 24	W19881	CEA-specific antibody	4.76e+02
34	41	60.3	123 24	W19889	CEA-specific antibody	4.76e+02
35	41	60.3	123 24	W19888	CEA-specific antibody	4.76e+02
36	41	60.3	123 24	W19887	CEA-specific antibody	4.76e+02
37	41	60.3	128 28	W31715	Gamma-1 heavy chain a	4.76e+02
38	41	60.3	268 27	W27136	Achromobacter lyticus	4.76e+02
39	41	60.3	268 27	W31403	Achromobacter lyticus	4.76e+02
40	41	60.3	268 26	W27135	Achromobacter lyticus	4.76e+02
41	41	60.3	397 19	W04270	B.t. alkaline protease	4.76e+02
42	41	60.3	481 5	R24442	Sequence of antibody	4.76e+02
43	41	60.3	512 39	W87797	Protease encoded by c	4.76e+02
44	41	60.3	512 14	R80505	S. lividans protease	4.76e+02
45	41	60.3	1170 12	R63231	Crystal protein CryET	4.76e+02

ALIGNMENTS

RESULT 1

ID W63157 standard; Protein; 216 AA.
AC W63157;
DE Human calcium channel beta-4 subunit.
KW Beta-4 subunit; human; calcium channel; assay; detection;
OS Characterisation; Lambert Eaton Syndrome; LES; diagnosis.
OS Homo sapiens.
PN US5792846-A.
PD 11-AUG-1998.
PF 31-MAY-1995; 455543.
PR 04-APR-1994; US-223305.
PR 04-APR-1988; US-176899.
PR 04-APR-1989; US-603751.
PR 04-APR-1989; WO-001408.
PR 20-FEB-1990; US-482384.
PR 30-NOV-1990; US-620250.
PR 15-AUG-1991; US-745206.
PR 31-MAY-1995; US-455543.
PA (SIBI-) SIBIA NEUROSCIENCES INC.
PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
PI Williams ME;
DR WPI; 98-456192/39.
DR N-PSDB; V42706.
PT DNA encoding human calcium channel alpha 1B subunit protein -
PT useful for recombinant production of the channel for screening of
PT its modulators, and diagnosis of Lambert Eaton Syndrome
PS Disclosure; Columns 225-228; 166pp; English.
CC The present sequence represents the beta-4 subunit of a human calcium
CC channel. Calcium channels are membrane-spanning, multi-subunit proteins
CC that allow controlled entry of calcium ions into cells. This leads
CC to depolarisation events required for muscle contraction. The recombinant
CC subunit, when expressed with nucleic acids encoding the complete calcium
CC channel, can be used in assays for the detection and characterisation of
CC compounds that modulate the channel. The DNA encoding the subunits can
CC be alternatively spliced when transcribed, giving more than one form of
CC the protein from the same transcript, each having slightly different
CC properties. In addition, the reactivity of the alpha 1 subunit with IgG
CC molecules from the serum of an individual with Lambert Eaton Syndrome
CC (LES) can be used as a diagnostic for the disease.


```

SQ Sequence 216 AA;
Query Match 79.4%; Score 54; DB 33; Length 216;
Best Local Similarity 75.0%; Pred. No. 1.78e+01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnfv 55
   |||:||||:|
QY 81 STSNDTTSAAVF 92

RESULT 2
ID R72605 standard; Protein; 216 AA.
AC R72605;
DE 01-DEC-1995 (first entry)
DT Human calcium channel subunit beta 1-4.
KW Calcium channel subunit; antagonist; agonist; diagnosis;
KW Lambert Eaton Syndrome.
OS Homo sapiens.
PN W09504822-A.
PD 16-FEB-1995.
PF 11-AUG-1994; U09230.
PR 11-AUG-1993; US-105536.
PR 05-NOV-1993; US-149097.
PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.
PI Ellis SB, Gillespie A, Harpold MM, McCue AF, Williams ME;
DR WPI: 95-090900/12.
DR N-PSDB: Q87832.
PT DNA encoding human calcium channel sub-unit(s) - used for
PT developing prods. for studying calcium channels, e.g. for
PT obtaining agonists and antagonists
PS Disclosure: Page 256-257; 285pp; English.
CC To isolate DNA encoding the Beta 1 subunit, a human hippocampus
CC cDNA library was screened by hybridisation to a DNA fragment
CC encoding a rabbit skeletal muscle calcium channel beta subunit.
CC A hybridising clone was selected and was in turn used to isolate
CC overlapping clones until the overlapping clones encompassing DNA
CC encoding the entire human calcium channel beta 2 subunit were
CC isolated and sequenced. Five alternatively spliced forms of the
CC beta 1 subunit have been identified. These forms are designated
CC beta 1-1, expressed in skeletal muscle, beta 1-2, expressed in the
CC CNS, beta 1-3, also expressed in the CNS, beta 1-4, expressed in
CC aorta tissue and HEK 293 cells, and beta 1-5, expressed in HEK 293
CC cells. Full-length DNA clones encoding the beta 1-2 and beta 1-3
CC subunits have been constructed. The subunits beta 1-1, beta 1-2,
CC beta 1-4 and beta 1-5 have been identified by nucleic acid
CC amplification analysis as alternatively spliced forms of the
CC beta subunit. Sequences of the beta 1 splice variants are set
CC forth in Q87838/R72611, Q87839/R72612 and Q87831-Q87833 and
CC R72604-R72706.
SQ Sequence 216 AA;

Query Match 79.4%; Score 54; DB 14; Length 219;
Best Local Similarity 75.0%; Pred. No. 1.78e+01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnfv 55
   |||:||||:|
QY 81 STSNDTTSAAVF 92

RESULT 4
ID W63158 standard; Protein; 219 AA.
AC W63158;
DT 12-OCT-1998 (first entry)
DE Human calcium channel beta-5 subunit.
KW Beta-5 subunit; human; calcium channel; assay; detection; ds.
KW Characterisation; Lambert Eaton Syndrome; LES; diagnosis; ds.
OS Homo sapiens.
PN US5792846-A.
PD 11-AUG-1998.
PF 31-MAY-1995; 455543.
PR 04-APR-1994; US-223305.
PR 04-APR-1988; US-176899.
PR 04-APR-1989; US-603751.
PR 04-APR-1989; WO-001408.
PR 20-FEB-1990; US-482384.
PR 30-NOV-1990; US-620250.
PR 15-AUG-1991; US-745206.
PR 31-MAY-1995; US-455543.
PA (STBI-) STBIA NEUROSCIENCES INC.
PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
PI Williams ME;
DR WPI: 98-456192/39.
DR N-PSDB: V42707.
PT DNA encoding human calcium channel alpha 1B subunit protein -
PT useful for recombinant production of the channel for screening of
PT its modulators, and diagnosis of Lambert Eaton Syndrome
PS Disclosure: Columns 227-230; 166pp; English.
CC The present sequence represents the beta-5 subunit of a human calcium
CC channel. Calcium channels are membrane-spanning, multi-subunit proteins
CC that allow controlled entry of calcium ions into cells. This leads
CC to depolarisation events required for muscle contraction. The recombinant
CC subunit, when expressed with nucleic acids encoding the complete calcium
CC channel, can be used in assays for the detection and characterisation of
CC compounds that modulate the channel. The DNA encoding the subunits can

```

```

PR 05-NOV-1993; US-149097.
PA (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.
PI Ellis SB, Gillespie A, Harpold MM, McCue AF, Williams ME;
DR WPI: 95-090900/12.
DR N-PSDB: Q87833.
PT DNA encoding human calcium channel sub-unit(s) - used for
PT developing prods. for studying calcium channels, e.g. for
PT obtaining agonists and antagonists
PS Disclosure: Page 258-259; 285pp; English.
CC To isolate DNA encoding the Beta 1 subunit, a human hippocampus
CC cDNA library was screened by hybridisation to a DNA fragment
CC encoding a rabbit skeletal muscle calcium channel beta subunit.
CC A hybridising clone was selected and was in turn used to isolate
CC overlapping clones until the overlapping clones encompassing DNA
CC encoding the entire human calcium channel beta 2 subunit were
CC isolated and sequenced. Five alternatively spliced forms of the
CC beta 1 subunit have been identified. These forms are designated
CC beta 1-1, expressed in skeletal muscle, beta 1-2, expressed in the
CC CNS, beta 1-3, also expressed in the CNS, beta 1-4, expressed in
CC aorta tissue and HEK 293 cells, and beta 1-5, expressed in HEK 293
CC cells. Full-length DNA clones encoding the beta 1-2 and beta 1-3
CC subunits have been constructed. The subunits beta 1-1, beta 1-2,
CC beta 1-4 and beta 1-5 have been identified by nucleic acid
CC amplification analysis as alternatively spliced forms of the
CC beta subunit. Sequences of the beta 1 splice variants are set
CC forth in Q87838/R72611, Q87839/R72612 and Q87831-Q87833 and
CC R72604-R72706.
SQ Sequence 219 AA;

Query Match 79.4%; Score 54; DB 14; Length 219;
Best Local Similarity 75.0%; Pred. No. 1.78e+01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnfv 55
   |||:||||:|
QY 81 STSNDTTSAAVF 92

RESULT 4
ID W63158 standard; Protein; 219 AA.
AC W63158;
DT 12-OCT-1998 (first entry)
DE Human calcium channel beta-5 subunit.
KW Beta-5 subunit; human; calcium channel; assay; detection; ds.
KW Characterisation; Lambert Eaton Syndrome; LES; diagnosis; ds.
OS Homo sapiens.
PN US5792846-A.
PD 11-AUG-1998.
PF 31-MAY-1995; 455543.
PR 04-APR-1994; US-223305.
PR 04-APR-1988; US-176899.
PR 04-APR-1989; US-603751.
PR 04-APR-1989; WO-001408.
PR 20-FEB-1990; US-482384.
PR 30-NOV-1990; US-620250.
PR 15-AUG-1991; US-745206.
PR 31-MAY-1995; US-455543.
PA (STBI-) STBIA NEUROSCIENCES INC.
PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
PI Williams ME;
DR WPI: 98-456192/39.
DR N-PSDB: V42707.
PT DNA encoding human calcium channel alpha 1B subunit protein -
PT useful for recombinant production of the channel for screening of
PT its modulators, and diagnosis of Lambert Eaton Syndrome
PS Disclosure: Columns 227-230; 166pp; English.
CC The present sequence represents the beta-5 subunit of a human calcium
CC channel. Calcium channels are membrane-spanning, multi-subunit proteins
CC that allow controlled entry of calcium ions into cells. This leads
CC to depolarisation events required for muscle contraction. The recombinant
CC subunit, when expressed with nucleic acids encoding the complete calcium
CC channel, can be used in assays for the detection and characterisation of
CC compounds that modulate the channel. The DNA encoding the subunits can

```

CC be alternatively spliced when transcribed, giving more than one form of
 CC the protein from the same transcript, each having slightly different
 CC properties. In addition, the reactivity of the alpha 1 subunit with IgG
 CC molecules from the serum of an individual with Lambert Eaton Syndrome
 CC (LES) can be used as a diagnostic for the disease.
 SQ Sequence 219 AA;

Query Match 79.4%; Score 54; DB 33; Length 219;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnsvf 55
 |||:||||: ||
 QY 81 STSNDTTSAAVF 92

RESULT 5

ID R39697 standard; Protein; 240 AA.
 AC R39697;
 DT 19-JAN-1994 (first entry)
 DE Myasthenic antigenic polypeptide.
 KW MAP; antibody detection; LEMS; Lambert-Eaton myasthenic syndrome;
 KW paraneoplastic sensory neuropathy; p mys B; mysB; ss.
 OS Homo sapiens.
 PN WO9314098-A.
 PD 22-JUL-1993.
 PF 11-JAN-1993; U00227.
 PR 10-JAN-1992; US-820312.
 PA (SLOK) SLOAN KETTERING INST CANCER.
 PI Furneaux HM, Posner JB;
 DR WPI: 93-243126/30.
 DR N-PSDB: Q46673.
 PT New purified myasthenic antigenic polypeptide and its corresp.
 PT antibody - useful for diagnosing and treating proliferation of
 PT neoplastic cells in patient with Lambert-Eaton myasthenic syndrome
 PS Claim 5; Page 26-27; 48pp; English.
 CC The sequence is that of myasthenic antigenic polypeptide (MAP)
 CC encoded by the cDNA clone p mysB. MAP can be used to detect
 CC antibodies associated with paraneoplastic sensory neuropathy
 CC such as Lambert-Eaton myasthenic syndrome (LEMS). These antibodies
 CC are used to determine if a patient with neurological symptoms has a
 CC tumour expressing MAP. To inhibit proliferation of neoplastic cells
 CC in patients with LEMS and for imaging neoplastic cells in LEMS
 CC patients.
 SQ Sequence 240 AA;

Query Match 79.4%; Score 54; DB 8; Length 240;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 17 stsdtttsnsvf 28
 |||:||||: ||
 QY 81 STSNDTTSAAVF 92

RESULT 6

ID R39563 standard; Protein; 478 AA.
 AC R39563;
 DT 09-FEB-1994 (first entry)
 DE Human neuronal VDCC beta-subunit encoded by clone HBBI.
 KW Voltage-dependent calcium channel; VDCC; beta-subunit;
 KW calcium flux; ss.
 OS Homo sapiens.
 PN DE422126-A.
 PD 19-AUG-1993.
 PF 06-JUL-1992; 222126.
 PR 17-FEB-1992; DE-204716.
 PR 06-JUL-1992; DE-222126.
 PA (FARB) BAYER AG.
 PI Spreyer P, Unterbeck A;
 DR WPI: 93-265734/34.
 DR N-PSDB: Q46076.
 PT Human neuronal beta-unit cDNA of voltage dependent calcium

PT channels - useful in calcium-flux studies and screening systems
 PT for agonists and antagonists of calcium channels
 PS Claim 2; Page 5-7; 13pp; German.
 CC A first oligonucleotide probe (Q46075) complementary to nucleotides
 CC 361-400 of the VDCC beta-subunit from rabbit skeletal muscle was
 CC used to screen a human hippocampus cDNA library. A 1.9kb cDNA
 CC fragment was isolated for further screening to isolate human VDCC
 CC beta-subunit coding sequences. Clone HBBI was sequenced and found
 CC to have 92% homology to the rabbit beta-subunit. The amino acid
 CC sequence R39563 was deduced from the open reading frame.
 CC See also Q46077-Q46078.
 SQ Sequence 478 AA;

Query Match 79.4%; Score 54; DB 8; Length 478;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnsvf 55
 |||:||||: ||
 QY 81 STSNDTTSAAVF 92

RESULT 7

ID W63143 standard; Protein; 478 AA.
 AC W63143;
 DT 12-OCT-1998 (first entry)
 DE Human calcium channel beta subunit splice variant beta2.
 KW Beta subunit; human; calcium channel; assay; detection;
 KW characterisation; Lambert Eaton Syndrome; LES; diagnosis.
 OS Homo sapiens.
 PN US5792846-A.
 PD 11-AUG-1998.
 PF 31-MAY-1995; 455543.
 PR 04-APR-1994; US-223305.
 PR 04-APR-1989; US-176899.
 PR 04-APR-1989; US-603751.
 PR 04-APR-1989; WO-U01408.
 PR 20-FEB-1990; US-482384.
 PR 30-NOV-1990; US-620250.
 PR 15-AUG-1991; US-745206.
 PR 31-MAY-1995; US-455543.
 PA (SIBI-) SIBIA NEUROSCIENCES INC.
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
 PI Williams ME;
 DR WPI: 98-456192/39.
 DR N-PSDB: V42687.

PT DNA encoding human calcium channel alpha 1B subunit protein -
 PT useful for recombinant production of the channel for screening of
 PT its modulators, and diagnosis of Lambert Eaton Syndrome
 PS Claim 3; Columns 217-220; 166pp; English.
 CC The present sequence is encoded by a splice variant of the beta subunit
 CC of a human calcium channel. Calcium channels are membrane-spanning,
 CC multi-subunit proteins that allow controlled entry of calcium ions into
 CC cells. This leads to depolarisation events required for muscle
 CC contraction. The recombinant subunit, when expressed with nucleic acids
 CC encoding the complete calcium channel, can be used in assays for the
 CC detection and characterisation of compounds that modulate the channel.
 CC The DNA encoding the subunits can be alternatively spliced when
 CC transcribed, giving more than one form of the protein from the same
 CC transcript, each having slightly different properties. In addition, the
 CC reactivity of the alpha 1 subunit with IgG molecules from the serum of
 CC an individual with Lambert Eaton Syndrome (LES) can be used as a
 CC diagnostic for the disease.
 SQ Sequence 478 AA;

Query Match 79.4%; Score 54; DB 33; Length 478;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnsvf 55
 |||:||||: ||
 QY 81 STSNDTTSAAVF 92

RESULT 8
 ID R72611 standard; Protein; 478 AA.
 AC R72611:
 DT 01-DEC-1995 (first entry)
 DE Human calcium channel subunit beta 1.
 KW Calcium channel subunit; antagonist; agonist; diagnosis;
 KW Lambert Eaton Syndrome.
 OS Homo sapiens.
 PN WO9504822-A.
 PD 16-FEB-1995.
 PF 11-AUG-1994; U09230.
 PR 11-AUG-1993; US-105536.
 PR 05-NOV-1993; US-149097.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Ellis SB, Gillespie A, Harpold MM, Mccue AF, Williams ME;
 DR WPI: 95-090900/12.
 DR N-PSDB: Q87838.
 PT DNA encoding human calcium channel sub-unit(s) - used for
 PT developing prods. for studying calcium channels, e.g. for
 PT obtaining agonists and antagonists
 PS Disclosure: Page 160-162; 285pp; English.
 CC To isolate DNA encoding the Beta 1 subunit, a human hippocampus
 CC cDNA library was screened by hybridisation to a DNA fragment
 CC encoding a rabbit skeletal muscle calcium channel beta subunit.
 CC A hybridising clone was selected and was in turn used to isolate
 CC overlapping clones until the overlapping clones encompassing DNA
 CC encoding the entire human calcium channel beta 2 subunit were
 CC isolated and sequenced. Five alternatively spliced forms of the
 CC beta 1 subunit have been identified. These forms are designated
 CC beta 1-1, expressed in skeletal muscle, beta 1-2, expressed in the
 CC CNS, beta 1-3, also expressed in the CNS, beta 1-4, expressed in
 CC aorta tissue and HEK 293 cells, and beta 1-5, expressed in HEK 293
 CC cells. Full-length DNA clones encoding the beta 1-2 and beta 1-3
 CC subunits have been constructed. The subunits beta 1-1, beta 1-2,
 CC beta 1-4 and beta 1-5 have been identified by nucleic acid
 CC amplification analysis as alternatively spliced forms of the
 CC beta subunit. Sequences of the beta 1 splice variants are set
 CC forth in Q87838/R72611, Q87839/R72612 and Q87831-Q87833 and
 CC R72604-R72706.
 SQ Sequence 478 AA;
 Query Match 79.4%; Score 54; DB 14; Length 478;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 44 stsdtttsnsvf 55
 QY 81 STSNDTTSAAVF 92
 III:IIII:II
 RESULT 9
 ID R33551 standard; Protein; 478 AA.
 AC R33551:
 DT 30-JUN-1993 (first entry)
 DE Sequence of splice variant beta 1-2 of beta human calcium
 DE channel subunit.
 KW Human calcium channel subunit; diagnosis; agonist; antagonist;
 KW Lambert Eaton syndrome.
 OS Homo sapiens.
 PN WO9304083-A.
 PD 04-MAR-1993.
 PF 14-AUG-1992; U06903.
 PR 15-AUG-1991; US-745206.
 PR 10-APR-1992; US-868354.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, Mccue AF,
 PI Williams ME;
 DR WPI: 93-093936/11.
 DR N-PSDB: Q37819.
 PT DNA encoding specific human calcium channel sub-units - used for
 PT identifying calcium channel agonists and antagonists and
 PT diagnosing Lambert Eaton syndrome

PS Disclosure: Page 129-131; 150pp; English.
 CC Five alternatively spliced forms of the human calcium channel
 CC beta 1 subunit have been identified and DNA encoding a number 1-1,
 CC of forms have been isolated. These forms are designated beta 1-1,
 CC expressed in skeletal muscle, beta 1-2, expressed in the CNS, beta
 CC 1-3, also expressed in the CNS, beta 1-4, expressed in aorta tissue
 CC and HEK 293 cells, and beta 1-5, expressed in HEK 293 cells.
 SQ Sequence 478 AA;
 Query Match 79.4%; Score 54; DB 6; Length 478;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 44 stsdtttsnsvf 55
 QY 81 STSNDTTSAAVF 92
 III:IIII:II
 RESULT 10
 ID R72604 standard; Protein; 479 AA.
 AC R72604:
 DT 01-DEC-1995 (first entry)
 DE Human calcium channel subunit beta 1-1.
 KW Calcium channel subunit; antagonist; agonist; diagnosis;
 KW Lambert Eaton Syndrome.
 OS Homo sapiens.
 PN WO9504822-A.
 PD 16-FEB-1995.
 PF 11-AUG-1994; U09230.
 PR 11-AUG-1993; US-105536.
 PR 05-NOV-1993; US-149097.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Ellis SB, Gillespie A, Harpold MM, Mccue AF, Williams ME;
 DR WPI: 95-090900/12.
 DR N-PSDB: Q87831.
 PT DNA encoding human calcium channel sub-unit(s) - used for
 PT developing prods. for studying calcium channels, e.g. for
 PT obtaining agonists and antagonists
 PS Disclosure: Page 253-256; 285pp; English.
 CC To isolate DNA encoding the Beta 1 subunit, a human hippocampus
 CC encoding a rabbit skeletal muscle calcium channel beta subunit.
 CC A hybridising clone was selected and was in turn used to isolate
 CC overlapping clones until the overlapping clones encompassing DNA
 CC encoding the entire human calcium channel beta 2 subunit were
 CC isolated and sequenced. Five alternatively spliced forms of the
 CC beta 1 subunit have been identified. These forms are designated
 CC beta 1-1, expressed in skeletal muscle, beta 1-2, expressed in the
 CC CNS, beta 1-3, also expressed in the CNS, beta 1-4, expressed in
 CC aorta tissue and HEK 293 cells, and beta 1-5, expressed in HEK 293
 CC cells. Full-length DNA clones encoding the beta 1-2 and beta 1-3
 CC subunits have been constructed. The subunits beta 1-1, beta 1-2,
 CC beta 1-4 and beta 1-5 have been identified by nucleic acid
 CC amplification analysis as alternatively spliced forms of the
 CC beta subunit. Sequences of the beta 1 splice variants are set
 CC forth in Q87838/R72611, Q87839/R72612 and Q87831-Q87833 and
 CC R72604-R72706.
 SQ Sequence 479 AA;
 Query Match 79.4%; Score 54; DB 14; Length 479;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 44 stsdtttsnsvf 55
 QY 81 STSNDTTSAAVF 92
 III:IIII:II
 RESULT 11
 ID W63156 standard; Protein; 523 AA.
 AC W63156:
 DT 12-OCT-1998 (first entry)
 DE Human calcium channel beta-1 subunit.

KW Beta-1 subunit; human; calcium channel; assay; detection;
 KW Characterisation; Lambert Eaton Syndrome; LES; diagnosis.
 OS Homo sapiens.
 PN US5792846-A.
 PD 11-AUG-1998. 455543.
 PF 31-MAY-1995; US-223305.
 PR 04-APR-1994; US-223305.
 PR 04-APR-1988; US-176899.
 PR 04-APR-1989; US-603751.
 PR 04-APR-1989; WO-U01408.
 PR 20-FEB-1990; US-482384.
 PR 30-NOV-1990; US-620250.
 PR 15-AUG-1991; US-745206.
 PR 31-MAY-1995; US-455543.
 PA (SIBI-) SIBIA NEUROSCIENCES INC.
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
 PI Williams ME;
 DR WPI: 98-456192/39.
 DR N-PSDB: V42705.
 PT DNA encoding human calcium channel alpha 1B subunit protein -
 PT useful for recombinant production of the channel for screening of
 PT its modulators, and diagnosis of Lambert Eaton Syndrome
 PS Claim 3: Columns 223-226; 166pp; English.
 CC The present sequence represents the beta-1 subunit of a human calcium
 CC channel. Calcium channels are membrane-spanning, multi-subunit proteins
 CC that allow controlled entry of calcium ions into cells. This leads
 CC to depolarisation events required for muscle contraction. The recombinant
 CC subunit, when expressed with nucleic acids encoding the complete calcium
 CC channel, can be used in assays for the detection and characterisation of
 CC compounds that modulate the channel. The DNA encoding the subunits can
 CC be alternatively spliced when transcribed, giving more than one form of
 CC the protein from the same transcript, each having slightly different
 CC properties. In addition, the reactivity of the alpha 1 subunit with IgG
 CC molecules from the serum of an individual with Lambert Eaton Syndrome
 CC (LES) can be used as a diagnostic for the disease.
 SQ Sequence 523 AA;

Query Match 79.4%; Score 54; DB 33; Length 523;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnfsv 55
 |||:||||: ||
 QY 81 STSNDTTSAAVF 92

RESULT 12
 ID W63147 standard; Protein: 530 AA;
 AC W63147;
 DT 12-OCT-1998 (first entry)
 DE Human calcium channel beta subunit.
 KW Beta subunit; human; calcium channel; assay; detection;
 KW Characterisation; Lambert Eaton Syndrome; LES; diagnosis.
 OS Homo sapiens.
 PN US5792846-A.
 PD 11-AUG-1998. 455543.
 PF 31-MAY-1995; US-223305.
 PR 04-APR-1994; US-223305.
 PR 04-APR-1988; US-176899.
 PR 04-APR-1989; US-603751.
 PR 04-APR-1989; WO-U01408.
 PR 20-FEB-1990; US-482384.
 PR 30-NOV-1990; US-620250.
 PR 15-AUG-1991; US-745206.
 PR 31-MAY-1995; US-455543.
 PA (SIBI-) SIBIA NEUROSCIENCES INC.
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
 PI Williams ME;
 DR WPI: 98-456192/39.
 DR N-PSDB: V42693.
 PT DNA encoding human calcium channel alpha 1B subunit protein -
 PT useful for recombinant production of the channel for screening of
 PT its modulators, and diagnosis of Lambert Eaton Syndrome

PS Disclosure; Columns 125-130; 166pp; English.
 CC The present sequence represents the beta subunit of a human calcium
 CC channel. Calcium channels are membrane-spanning, multi-subunit
 CC proteins that allow controlled entry of calcium ions into cells.
 CC This leads to depolarisation events required for muscle contraction.
 CC The recombinant subunit, when expressed with nucleic acids encoding
 CC the complete calcium channel, can be used in assays for the detection
 CC and characterisation of compounds that modulate the channel. The
 CC DNA encoding the subunits can be alternatively spliced when
 CC transcribed, giving more than one form of the protein from the same
 CC transcript, each having slightly different properties. In addition, the
 CC reactivity of the alpha 1 subunit with IgG molecules from the serum of
 CC an individual with Lambert Eaton Syndrome (LES) can be used as a
 CC diagnostic for the disease.
 SQ Sequence 530 AA;

Query Match 79.4%; Score 54; DB 33; Length 530;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnfsv 55
 |||:||||: ||
 QY 81 STSNDTTSAAVF 92

RESULT 13
 ID R39565 standard; Protein: 571 AA.
 AC R39565;
 DT 09-FEB-1994 (first entry)
 DE Human neuronal VDCC beta-subunit encoded by clone HBB3.
 KW Voltage-dependent calcium channel; VDCC; beta-subunit;
 KW calcium flux; ss.
 OS Homo sapiens.
 PN DE422126-A.
 PD 19-AUG-1993. 222126.
 PF 06-JUL-1992; 222126.
 PR 17-FEB-1992; DE-204716.
 PR 06-JUL-1992; DE-222126.
 PA (FARB) BAYER AG.
 PI Spreyer P, Unterbeck A;
 DR WPI: 93-265734/34.
 DR N-PSDB: Q46078.
 PT Human neuronal beta-unit cDNA of voltage dependent calcium
 PT channels - useful in calcium-flux studies and screening systems
 PT for agonists and antagonists of calcium channels
 PS Claim 2: Page 11-13; 13pp; German.
 CC A first oligonucleotide probe (Q46075) complementary to nucleotides
 CC 361-400 of the VDCC beta-subunit from rabbit skeletal muscle was
 CC used to screen a human hippocampus cDNA library. A 1.9kb cDNA
 CC fragment was isolated for further screening to isolate human VDCC
 CC beta-subunit coding sequences. Clone HBB3 was sequenced and
 CC nucleotides 1-1288 were found to have 92% homology to the rabbit
 CC beta-subunit; from position 1289 there is no detectable homology.
 CC The amino acid sequence R39565 was deduced from the open reading
 CC frame. See also Q46076 and Q46077.
 SQ Sequence 571 AA;

Query Match 79.4%; Score 54; DB 8; Length 571;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnfsv 55
 |||:||||: ||
 QY 81 STSNDTTSAAVF 92

RESULT 14
 ID W63144 standard; Protein: 598 AA.
 AC W63144;
 DT 12-OCT-1998 (first entry)
 DE Human calcium channel beta subunit splice variant beta3.
 KW Beta subunit; human; calcium channel; assay; detection;
 KW Characterisation; Lambert Eaton Syndrome; LES; diagnosis.

OS Homo sapiens.
 PN US5792846-A.
 PD 11-AUG-1998.
 PF 31-MAY-1995; 455543.
 PR 04-APR-1994; US-223305.
 PR 04-APR-1988; US-176899.
 PR 04-APR-1989; US-603753.
 PR 04-APR-1989; WO-U01408.
 PR 20-FEB-1990; US-482384.
 PR 30-NOV-1990; US-620250.
 PR 15-AUG-1991; US-745206.
 PR 31-MAY-1995; US-455543.
 PA (SIBI-) SIBIA NEUROSCIENCES INC.
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
 PI Williams ME;
 DR WPI: 98-456192/39.
 DR N-PSDB; V42688.
 DR DNA encoding human calcium channel alpha 1B subunit protein -
 PT useful for recombinant production of the channel for screening of
 PT its modulators, and diagnosis of Lambert Eaton Syndrome
 PS Claim 3: Columns 219-224; 166pp; English.
 CC The present sequence represents a splice variant of the beta subunit of
 CC a human calcium channel. Calcium channels are membrane-spanning,
 CC multi-subunit proteins that allow controlled entry of calcium ions into
 CC cells. This leads to depolarisation events required for muscle
 CC contraction. The recombinant subunit, when expressed with nucleic acids
 CC encoding the complete calcium channel, can be used in assays for the
 CC detection and characterisation of compounds that modulate the channel.
 CC The DNA encoding the subunits can be alternatively spliced when
 CC transcribed, giving more than one form of the protein from the same
 CC transcript, each having slightly different properties. In addition, the
 CC reactivity of the alpha 1 subunit with IgG molecules from the serum of
 CC an individual with Lambert Eaton Syndrome (LES) can be used as a
 CC diagnostic for the disease.
 SQ Sequence 598 AA;

Query Match 79.4%; Score 54; DB 33; Length 598;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnfv 55
 |||:||||:|
 QY 81 STSNDTTSAAVF 92

RESULT 15
 ID R72612 standard; Protein; 598 AA.
 AC R72612;
 DT 01-DEC-1995 (first entry)
 DE Human calcium channel subunit beta 1-3.
 KW Calcium channel subunit; antagonist; agonist; diagnosis;
 KW Lambert Eaton Syndrome.
 OS Homo sapiens.
 PN W09504822-A.
 PD 16-FEB-1995.
 PF 11-AUG-1994; U09230.
 PR 11-AUG-1993; US-105536.
 PR 05-NOV-1993; US-149097.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Ellis SB, Gillespie A, Harpold MM, McCue AF, Williams ME;
 DR WPI: 95-090900/12.
 DR N-PSDB; Q87839.
 PT DNA encoding human calcium channel sub-unit(s) - used for
 PT developing prods. for studying calcium channels, e.g. for
 PT obtaining agonists and antagonists
 PS Disclosure; Page 163-165; 285pp; English.
 CC To isolate DNA encoding the Beta 1 subunit, a human hippocampus
 CC cDNA library was screened by hybridisation to a DNA fragment
 CC encoding a rabbit skeletal muscle calcium channel beta subunit.
 CC A hybridising clone was selected and was in turn used to isolate
 CC overlapping clones until the overlapping clones encompassing DNA
 CC encoding the entire human calcium channel beta 2 subunit were
 CC isolated and sequenced. Five alternatively spliced forms of the

CC beta 1 subunit have been identified. These forms are designated
 CC beta 1-1, expressed in skeletal muscle, beta 1-2, expressed in the
 CC CNS, beta 1-3, also expressed in the CNS, beta 1-4, expressed in
 CC aorta tissue and HEK 293 cells, and beta 1-5, expressed in HEK 293
 CC cells. Full-length DNA clones encoding the beta 1-2 and beta 1-3
 CC subunits have been constructed. The subunits beta 1-1, beta 1-2,
 CC beta 1-4 and beta 1-5 have been identified by nucleic acid
 CC amplification analysis as alternatively spliced forms of the
 CC beta subunit. Sequences of the beta 1 splice variants are set
 CC forth in Q87838/R72611, Q87839/R72612 and Q87831-Q87833 and
 CC R72604-R72706.
 SQ Sequence 598 AA;

Query Match 79.4%; Score 54; DB 14; Length 598;
 Best Local Similarity 75.0%; Pred. No. 1.78e+01;
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 stsdtttsnfv 55
 |||:||||:|
 QY 81 STSNDTTSAAVF 92

Search completed: Wed Sep 1 16:33:43 1999
 Job time : 36 secs.


```

RESULT 2
ENTRY
TITLE L-type voltage-gated calcium channel B subunit - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
12-Aug-1996
ACCESSIONS I65766 #type complete
REFERENCE I52859 L-type voltage-gated calcium channel B subunit - human
#authors Collin, T.; Wang, J.; Nargeot, J.; Schwartz, A.
#journal Circ. Res. (1993) 72:1337-1344
#title Molecular cloning of three isoforms of the L-type
voltage-dependent calcium channel B subunit from normal
human heart.
#cross-references MUID:93265672
#accession I65766 preliminary; translated from GB/EMBL/DBJ
#status #preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-477 #label RES
#cross-references GB:L06111; NID:gl87016; PID:gl87017
SUMMARY #length 477 #molecular_weight 52977 #checksum 4957

Query Match 79.4%; Score 54; DB 2; Length 477;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 43 STSDTTNSFV 54
|||:||||:|
QY 81 STSNDTTSAFV 92

RESULT 3
ENTRY
TITLE JH0566 #type complete
ORGANISM calcium channel beta-2 chain - human
DATE #formal_name Homo sapiens #common_name man
30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
10-Sep-1997
ACCESSIONS JH0566
REFERENCE JH0564
#authors Williams, M.E.; Feldman, D.H.; McCue, A.F.; Brenner, R.;
Velicelebi, G.; Ellis, S.B.; Harpold, M.M.
#journal Neuron (1992) 8:71-84
#title Structure and functional expression of alpha1, alpha2, and
beta subunits of a novel human neuronal calcium channel
subtype.
#cross-references MUID:92110010
#accession JH0566
#molecule_type mRNA
#residues 1-478 #label WIL
#cross-references GB:M76560; NID:gl79743; PID:gl79744
#experimental_source hippocampus
COMMENT This protein is a subunit of the voltage dependent calcium channel.
KEYWORDS glycoprotein; phosphoprotein
FEATURE
32,167,209,348,374, #binding_site phosphate (Ser) (covalent) (by protein
450,464 kinase C) #status predicted\
64,201 #binding_site phosphate (Thr) (covalent) (by protein
kinase C) #status predicted\
189,425 #binding_site carbohydrate (Asn) (covalent) #status
predicted\
205 #binding_site phosphate (Thr) (covalent) (by
CAMP-dependent kinase) #status predicted
SUMMARY #length 478 #molecular_weight 52934 #checksum 6465

Query Match 79.4%; Score 54; DB 2; Length 478;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTNSFV 55
|||:||||:|
QY 81 STSNDTTSAFV 92

RESULT 4
ENTRY
TITLE B44461 #type complete
ORGANISM voltage-dependent calcium channel beta subunit beta1B1
DATE #formal_name Homo sapiens #common_name man
10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change
20-Mar-1998
ACCESSIONS B44461
REFERENCE A44461
#authors Powers, P.A.; Liu, S.; Hogan, K.; Gregg, R.G.
#journal J. Biol. Chem. (1992) 267:22967-22972
#title Skeletal muscle and brain isoforms of a beta-subunit of human
voltage-dependent calcium channels are encoded by a single
gene.
#cross-references MUID:93054616
#accession B44461 preliminary
#status #preliminary
#molecule_type nucleic acid
#residues 1-478 #label POW
#cross-references GB:M92302; NID:gl79803; PID:gl79804
#experimental_source hippocampus
#note sequence inconsistent with the nucleotide translation
sequence extracted from NCBI backbone (NCBIP:118131)
#note #length 478 #molecular_weight 53176 #checksum 7630
SUMMARY

Query Match 79.4%; Score 54; DB 2; Length 478;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTNSFV 55
|||:||||:|
QY 81 STSNDTTSAFV 92

RESULT 5
ENTRY
TITLE I65767 #type complete
ORGANISM L-type voltage-gated calcium channel B subunit - human
DATE #formal_name Homo sapiens #common_name man
12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
12-Aug-1996
ACCESSIONS I65767
REFERENCE I52859
#authors Collin, T.; Wang, J.; Nargeot, J.; Schwartz, A.
#journal Circ. Res. (1993) 72:1337-1344
#title Molecular cloning of three isoforms of the L-type
voltage-dependent calcium channel B subunit from normal
human heart.
#cross-references MUID:93265672
#accession I65767 preliminary; translated from GB/EMBL/DBJ
#status #preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-522 #label RES
#cross-references GB:L06112; NID:gl87018; PID:gl87019
SUMMARY #length 522 #molecular_weight 57834 #checksum 7462

Query Match 79.4%; Score 54; DB 2; Length 522;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 43 STSDTTNSFV 54
|||:||||:|
QY 81 STSNDTTSAFV 92

RESULT 6
ENTRY
TITLE A41347 #type complete
ORGANISM calcium channel protein beta chain,
dihydropyridine-sensitive, skeletal muscle - rabbit
DATE #formal_name Oryctolagus cuniculus #common_name domestic
rabbit
03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change
07-Feb-1997
ACCESSIONS A41347

```

```

REFERENCE      A41347
#authors      Ruth, P.; Roehrkasten, A.; Biel, M.; Bosse, E.; Regulla, S.;
#journal      Meyer, H.E.; Flockerzi, V.; Hofmann, F.
#title       Science (1989) 245:1115-1118
#cross-references GB:M92303; NID:g291880; PID:g179806
#accession    A41347
#status      preliminary
#molecule_type mRNA
#residues     1-524 #label RUT
#cross-references GB:M25817
#keywords     skeletal muscle
#note        length 524 #molecular-weight 57868 #checksum 8438
SUMMARY
Query Match      79.4%; Score 54; DB 2; Length 524;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTNSFV 55
   |||:||||:|
Qy 81 STSNDTTSAFV 92

RESULT 7
ENTRY   A48895 #type complete
TITLE   myasthenic syndrome antigen B - human
ALTERNATE_NAMES
ORGANISM MYSB
#formal_name Homo sapiens #common_name man
DATE     07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change
20-Mar-1998
ACCESSIONS
REFERENCE A48895
#authors  Rosenfeld, M.R.; Wong, E.; Dalmau, J.; Manley, G.; Posner,
#journal  J.B.; Sher, E.; Furneaux, H.M.
#title   Ann. Neurol. (1993) 33:113-120
#cross-references MUID:93263585
#accession A48895
#status    preliminary
#residues  1-567 #label ROS
#cross-references GB:S60415; NID:g300416; PID:g300417
#experimental_source fetal brain
#note      sequence extracted from NCBI backbone (NCBIN:132135,
          NCBIP:132136)
KEYWORDS  alternative splicing
SUMMARY   length 567 #molecular-weight 62087 #checksum 9649
Query Match      79.4%; Score 54; DB 2; Length 567;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 58 STSDTTNSFV 69
   |||:||||:|
Qy 81 STSNDTTSAFV 92

RESULT 8
ENTRY   C44461 #type complete
TITLE   voltage-dependent calcium channel beta subunit beta1B2
ORGANISM isoform - human
#formal_name Homo sapiens #common_name man
DATE     10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change
20-Mar-1998
ACCESSIONS
REFERENCE C44461
#authors  Powers, P.A.; Liu, S.; Hogan, K.; Gregg, R.G.
#journal  J. Biol. Chem. (1992) 267:22967-22972
#title   Skeletal muscle and brain isoforms of a beta-subunit of human
          voltage-dependent calcium channels are encoded by a single
          gene.

```

```

#cross-references MUID:93054616
#accession    C44461
#status      preliminary; not compared with conceptual translation
#molecule_type nucleic acid
#residues     1-596 #label POW
#cross-references GB:M92303; NID:g291880; PID:g179806
#experimental_source hippocampus
#note        sequence extracted from NCBI backbone (NCBIP:118133)
          length 596 #molecular-weight 65808 #checksum 4540
SUMMARY
Query Match      79.4%; Score 54; DB 2; Length 596;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTNSFV 55
   |||:||||:|
Qy 81 STSNDTTSAFV 92

RESULT 9
ENTRY   S18304 #type complete
TITLE   calcium channel protein beta chain - rat
ALTERNATE_NAMES
ORGANISM Rattus norvegicus #common_name Norway rat
DATE     22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
20-Mar-1998
ACCESSIONS
REFERENCE S18304
#authors  Pragnell, M.; Sakamoto, J.; Jay, S.D.; Campbell, K.P.
#journal  FEBS Lett. (1991) 291:253-258
#title   Cloning and tissue-specific expression of the brain calcium
          channel beta-subunit.
#cross-references MUID:92038046
#accession S18304
#molecule_type mRNA
#residues     1-597 #label PRA
#cross-references GB:X61394; NID:g55893; PID:g55894
#note      length 597 #molecular-weight 65679 #checksum 8572
SUMMARY
Query Match      79.4%; Score 54; DB 2; Length 597;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTNSFV 55
   |||:||||:|
Qy 81 STSNDTTSAFV 92

RESULT 10
ENTRY   I52859 #type complete
TITLE   L-type voltage-gated calcium channel B subunit - human
ALTERNATE_NAMES
ORGANISM Homo sapiens #common_name man
DATE     02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
02-Jul-1996
ACCESSIONS
REFERENCE I52859
#authors  Collin, T.; Wang, J.; Nargeot, J.; Schwartz, A.
#journal  Circ. Res. (1993) 72:1337-1344
#title   Molecular cloning of three isoforms of the L-type
          voltage-dependent calcium channel B subunit from normal
          human heart.
#cross-references MUID:93265672
#accession I52859
#status      preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues     1-597 #label RES
#cross-references GB:L06110; NID:g187014; PID:g187015
#note      length 597 #molecular-weight 65578 #checksum 5631
SUMMARY
Query Match      79.4%; Score 54; DB 2; Length 597;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTNSFV 55

```



```
||||:||||:|
QY 81 STSNDTTSAAFY 92

RESULT 11
ENTRY S21048 #type complete
TITLE calcium channel protein beta chain CaB2b - rabbit
ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
        rabbit
DATE 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change
        10-Sep-1997
ACCESSIONS S21048
REFERENCE S21046
#authors Hullin, R.; Singer-Iahat, D.; Freichel, M.; Biel, M.; Dascal,
        N.; Hofmann, F.; Flockerzi, V.
#journal EMBO J. (1992) 11:885-890
#title Calcium channel beta subunit heterogeneity: functional
        expression of cloned cDNA from heart, aorta and brain.
#cross-references MIM:92192022
#accession S21048
#molecule_type mRNA
#residues 1-632 #label HUL
#cross-references EMBL:X64298; NID:g1499; PID:g1500
SUMMARY #length 632 #molecular-weight 70943 #checksum 1583

Query Match 79.4%; Score 54; DB 2; Length 632;
Best Local Similarity 75.0%; Pred. No. 1.44e-01;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 30 STSSDTSNSFV 41
||||:||||:|
QY 81 STSNDTTSAAFY 92

RESULT 12
ENTRY D56271 #type complete
TITLE long polar fibrillar operon protein LpfD - Salmonella
        typhimurium
ORGANISM #formal_name Salmonella typhimurium
DATE 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change
        09-Sep-1997
ACCESSIONS D56271
REFERENCE A56271
#authors Baecumler, A.J.; Heffron, F.
#journal J. Bacteriol. (1995) 177:2087-2097
#title Identification and sequence analysis of lpfABCD, a putative
        fibrillar operon of Salmonella typhimurium.
#accession D56271
#status preliminary
#molecule_type DNA
#residues 1-355 #label BAE
#cross-references GB:U18559; NID:g829370; PID:g829374
GENETICS lpfD
#gene
SUMMARY #length 355 #molecular-weight 37714 #checksum 5448

Query Match 70.6%; Score 48; DB 2; Length 355;
Best Local Similarity 63.6%; Pred. No. 3.61e+00;
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 51 TSKNTTGATFV 61
||||:||||:|
QY 82 TSNDTTSAAFY 92

RESULT 13
ENTRY A57172 #type complete
TITLE probable hormone receptor EMRI precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change
        24-Sep-1998
ACCESSIONS A57172
REFERENCE A57172
#authors
```

```
#authors Baud, V.; Chisoe, S.L.; Viegas-Pequignot, E.; Diriong, S.;
        N'Guyen, V.C.; Roe, B.A.; Lipinski, M.
#journal Genomics (1995) 26:334-344
#title EMRI, an unusual member in the family of hormone receptors
        with seven transmembrane segments.
#accession A57172
#status preliminary
#molecule_type mRNA
#residues 1-886 #label BAU
#cross-references GB:X81479; NID:g784993; PID:g784994
GENETICS
#gene GDB:EMRI
#cross-references GDB:378349; OMIM:600493
#map_position 19p13.3-19p13.3
KEYWORDS transmembrane protein
SUMMARY #length 886 #molecular-weight 97679 #checksum 2055

Query Match 70.6%; Score 48; DB 2; Length 886;
Best Local Similarity 61.5%; Pred. No. 3.61e+00;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 470 SESTETTGAFVS 482
||||:||||:|
QY 81 STSNDTTSAAFY 93

RESULT 14
ENTRY A34755 #type complete
TITLE nitrogen regulatory protein nit-2 - Neurospora crassa
ORGANISM #formal_name Neurospora crassa
DATE 13-Jul-1990 #sequence_revision 26-Jul-1996 #text_change
        16-Feb-1997
ACCESSIONS A34755
REFERENCE A34755
#authors Fu, Y.H.; Marzluf, G.A.
#journal Mol. Cell. Biol. (1990) 10:1056-1065
#title nit-2, the major nitrogen regulatory gene of Neurospora
        crassa, encodes a protein with a putative zinc finger
        DNA-binding domain.
#cross-references MIM:90158568
#accession A34755
#molecule_type DNA; mRNA
#residues 1-1036 #label FUY
#cross-references GB:M33956
GENETICS
#introns 209/2; 335/3
CLASSIFICATION #superfamily nitrogen regulatory protein nit-2; GATA-type
        zinc finger homology
KEYWORDS DNA binding; transcription regulation; zinc finger
FEATURE 740-793
        743-767,
SUMMARY #length 1036 #molecular-weight 109295 #checksum 235

Query Match 70.6%; Score 48; DB 1; Length 1036;
Best Local Similarity 53.8%; Pred. No. 3.61e+00;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 87 SSSNNNTSNGFVA 99
||||:||||:|
QY 81 STSNDTTSAAFY 93

RESULT 15
ENTRY S50536 #type complete
TITLE hypothetical protein YER033c - yeast (Saccharomyces
        cerevisiae)
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 28-May-1993 #sequence_revision 24-Feb-1995 #text_change
        21-Nov-1997
ACCESSIONS S50536
REFERENCE S50428
#authors Dietrich, F.S.
```

#submission submitted to the EMBL Data Library, December 1994
#description The sequence of S. cerevisiae cosmids 9379, 9581, and lambda
clone 4678.

#accession S50536

##molecule_type DNA

##residues_type 1-1076 ##label DIE

##cross-references EMBL:U18796; NID:g603265; PID:g603266; MIPS:YER033c

GENETICS

#map_position 5R

SUMMARY #length 1076 #molecular-weight 119349 #checksum 9442

Query Match

Best Local Similarity 67.6%; Score 46; DB 2; Length 1076;

Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 97 TSTNQTTSNSFV 108

QY 81 STSNDTTSAAFV 92

...:|:|:|:|

Search completed: Wed Sep 1 16:32:49 1999
Job time : 14 secs.

 W O S R L H (TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed Sep 1 16:31:14 1999; MasPar time 2.98 Seconds
 Tabular output not generated.
 123.152 Million cell updates/sec

Title: >PCT-US99-13024-2
 Description: (81-93) from PCTUS9913024.pep (12 of 12)
 Perfect Score: 68
 Sequence: 1 STSNDTTSAAAFVS 13

Scoring table: PAM 150
 Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot37
 1:swissprot

Statistics: Mean 22.697; Variance 22.190; scale 1.023

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	54	79.4	478	1	CCBB_HUMAN	DIHYDROPYRIDINE-SENSIT	3.03e-02
2	54	79.4	523	1	CCBC_HUMAN	DIHYDROPYRIDINE-SENSIT	3.03e-02
3	54	79.4	524	1	CCBC_RABIT	DIHYDROPYRIDINE-SENSIT	3.03e-02
4	54	79.4	567	1	CCB2_HUMAN	DIHYDROPYRIDINE-SENSIT	3.03e-02
5	54	79.4	596	1	CCBA_HUMAN	DIHYDROPYRIDINE-SENSIT	3.03e-02
6	54	79.4	597	1	CCBA_RAT	DIHYDROPYRIDINE-SENSIT	3.03e-02
7	54	79.4	632	1	CCB2_RABIT	DIHYDROPYRIDINE-SENSIT	3.03e-02
8	48	70.6	355	1	LPFD_SALTY	LPFD PROTEIN PRECURSOR	1.11e+00
9	48	70.6	886	1	EMRL_HUMAN	CELL SURFACE GLYCOPROT	1.11e+00
10	48	70.6	1036	1	NIT2_NEUCR	NITROGEN CATABOLIC ENZ	1.11e+00
11	46	67.6	532	1	CPPI_DROME	CYTOCHROME P450 12B1 P	1.11e+00
12	46	67.6	1076	1	YEM3_YEAST	HYPOTHETICAL 119.3 KD	3.44e+00
13	45	66.2	345	1	PURA_METJA	ADENYLOSUCCLINATE SYNTH	5.97e+00
14	45	66.2	416	1	IR12_HCMVA	HYPOTHETICAL PROTEIN 1	5.97e+00
15	45	66.2	3066	1	POLG_BCMVN	GENOME POLYPROTEIN [CO	5.97e+00
16	44	64.7	490	1	C1KL_DROME	VOLTAGE-GATED POTASSIU	1.03e+01
17	44	64.7	588	1	PR28_YEAST	PRE-MRNA SPLICING FACT	1.03e+01
18	44	64.7	607	1	HRA1_XANCV	HYPERSENSITIVITY RESPO	1.03e+01
19	44	64.7	3106	1	LMA2_MOUSE	LAMININ ALPHA-2 CHAIN	1.03e+01
20	44	64.7	3110	1	LMA2_HUMAN	LAMININ ALPHA-2 CHAIN	1.03e+01
21	44	63.2	142	1	RPC9_YEAST	DNA-DIRECTED RNA POLYM	1.75e+01
22	43	63.2	269	1	FOLD_MYCPN	METHYLENETETRAHYDROFOL	1.75e+01
23	43	63.2	337	1	YEB7_YEAST	HYPOTHETICAL 38.2 KD P	1.75e+01

24 43 63.2 477 1 AMT_MYCTU PROBABLE AMMONIUM TRAN 1.75e+01
 25 43 63.2 500 1 CACT_DROME DEVELOPMENTAL PROTEIN 1.75e+01
 26 43 63.2 573 1 YEB8_YEAST HYPOTHETICAL 63.7 KD P 1.75e+01
 27 43 63.2 757 1 PIGR_BOVIN POLYMERIC-IMMUNOGLOBUL 1.75e+01
 28 43 63.2 786 1 YS8A_CAEEL HYPOTHETICAL 84.3 KD P 1.75e+01
 29 43 63.2 1102 1 YK64_CAEEL HYPOTHETICAL 124.8 KD 1.75e+01
 30 43 63.2 1124 1 TRPL_DROME TRANSPARENT-RECEPTOR-POT 1.75e+01
 31 43 63.2 1234 1 JMJ_MOUSE JUMONJI PROTEIN. 1.75e+01
 32 42 61.8 290 1 YEAB_BACSU HYPOTHETICAL 31.8 KD P 2.94e+01
 33 42 61.8 309 1 YF44_METH HYPOTHETICAL SUGAR KIN 2.94e+01
 34 42 61.8 325 1 YHU3_YEAST HYPOTHETICAL 33.4 KD P 2.94e+01
 35 42 61.8 351 1 PERX_SOLTU SUBERIZATION-ASSOCIATE 2.94e+01
 36 42 61.8 363 1 PER2_LYCES SUBERIZATION-ASSOCIATE 2.94e+01
 37 42 61.8 364 1 PER1_LYCES SUBERIZATION-ASSOCIATE 2.94e+01
 38 42 61.8 429 1 HISX_METJA HISTIDINOL DEHYDROGENA 2.94e+01
 39 42 61.8 457 1 TBB1_PORPU TUBULIN BETA CHAIN 1. 2.94e+01
 40 42 61.8 514 1 AMT2_LYCES HIGH AFFINITY AMMONIUM 2.94e+01
 41 42 61.8 871 1 YAY7_SCHPO HYPOTHETICAL 98.4 KD P 2.94e+01
 42 42 61.8 887 1 YAY5_SCHPO PROBABLE ATP-DEPENDENT 2.94e+01
 43 42 61.8 953 1 PSD1_HUMAN 26S PROTEASOME REGULAT 2.94e+01
 44 42 61.8 1333 1 YN99_YEAST PROBABLE ATP-DEPENDENT 2.94e+01
 45 42 61.8 1475 1 N153_HUMAN NUCLEAR PORE COMPLEX P 2.94e+01

ALIGNMENTS

RESULT 1
 ID CCBB_HUMAN STANDARD; PRT; 478 AA.
 AC Q02640;
 DT 01-JUN-1994 (REL. 29, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CALCIUM CHANNEL BETA-1-B1 SUBUNIT
 DE (BETA-1 ISOFORM B) (BETA-2) (BETA-1C).
 GN CACNB1 OR CACNLB1.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-HIPPOCAMPUS;
 RA POWERS P.A., LIU S., HOGAN K., GREGG R.G.;
 RT "Skeletal muscle and brain isoforms of a beta-subunit of human
 voltage-dependent calcium channels are encoded by a single gene.";
 RL J. BIOL. CHEM. 267:22967-22972(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE; 92110010.
 RA WILLIAMS M.E., FELDMAN D.H., MCCUE A.F., BRENNER R.,
 RA VELICELEBI G., ELLIS S.B., HARPOLD M.M.;
 RT "Structure and functional expression of alpha 1, alpha 2, and beta
 subunits of a novel human neuronal calcium channel subtype.";
 RL NEURON 8:71-84(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-HEART;
 RA COLLIN T., WANG J., NARGEOT J., SCHWARTZ A.;
 RT "Molecular cloning of three isoforms of the L-type voltage-dependent
 calcium channel beta subunit from normal human heart.";
 RL CIRC. RES. 72:1337-1344(1993).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA POWERS P.A., GREGG R.G., HOGAN K.;
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -!- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROPYRIDINE (DHP)
 CC SENSITIVE CALCIUM CHANNEL.
 CC -!- SURUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:
 CC ALPHA-1, ALPHA-2, BETA AND GAMMA.
 CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ASSOCIATED WITH
 CC THE CYTOPLASMIC ASPECT OF THE SARCOLEMMMA.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, HEART AND SPLEEN.

```

CC -!- ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1
CC SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M92302; G179804; -
DR EMBL: M76560; G179744; -
DR EMBL: L06111; G187017; -
DR EMBL: U86960; G2155254; JOINED.
DR EMBL: U86952; G2155254; JOINED.
DR EMBL: U86953; G2155254; JOINED.
DR EMBL: U86954; G2155254; JOINED.
DR EMBL: U86955; G2155254; JOINED.
DR EMBL: U86956; G2155254; JOINED.
DR EMBL: U86957; G2155254; JOINED.
DR EMBL: U86958; G2155254; JOINED.
DR EMBL: U86959; G2155254; JOINED.
DR PIR: B44461; B44461.
DR MIM: I14207; -
DR PROSITE: PS50002; SH3; 1.
DR PFAM: PF00018; SH3; 1.
DR PFAM: PF00774; Ca_channel_B; 1.
KW IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;
KW GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;
KW MULTIGENE FAMILY.
FT DOMAIN 100 161 SH3.
FT CARBOHYD 189 189 POTENTIAL.
FT CARBOHYD 425 425 POTENTIAL.
FT CONFLICT 21 21 E -> G (IN REF. 2).
FT CONFLICT 28 29 QG -> R (IN REF. 3).
FT CONFLICT 434 435 AA -> RR (IN REF. 1 AND 4).
SQ SEQUENCE 478 AA; 53006 MW; 6011892D CRC32;

Query Match 79.48; Score 54; DB 1; Length 478;
Best Local Similarity 75.08; Pred. No. 3.03e-02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTSNSFV 55
QY 81 STSDTTSRAFV 92
|||||:|:|
|:|:|:|:|:|

RESULT 2
ID CCBC_HUMAN STANDARD; PRT; 523 AA.
AC Q02639;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CALCIUM CHANNEL BETA-1M SUBUNIT
DE (BETA-1 ISOFORM C) (BETA-1A).
GN CACNB1 OR CACNLB1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SKELETAL MUSCLE;
RX MEDLINE: 93054616.
RA POWERS P.A., LIU S., HOGAN K., GREGG R.G.;
RT "Skeletal muscle and brain isoforms of a beta-subunit of human
RT voltage-dependent calcium channels are encoded by a single gene.";
RL J. BIOL. CHEM. 267:22967-22972(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-HEART;

```

```

RX MEDLINE: 93265672.
RA COLLIN T., WANG J., NARGEOT J., SCHWARTZ A.;
RT "Molecular cloning of three isoforms of the L-type voltage-dependent
RT calcium channel beta subunit from normal human heart.";
RL CIRC. RES. 72:1337-1344(1993).
RN [3]
RP SEQUENCE FROM N.A.
RA POWERS P.A., GREGG R.G., HOGAN K.;
RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE OF 146-209 FROM N.A.
RX MEDLINE: 93372845.
RA ILES D.E., SEGERS B., SENGERS R.C.A., MONSIEURS K., HEYTENS L.,
RA HALSALL P.J., HOPKINS P.M., ELLIS F.R., HALL-CURRAN J.L.,
RA STEWART A.D., WIERINGA B.;
RT "Genetic mapping of the beta 1- and gamma-subunits of the human
RT skeletal muscle L-type voltage-dependent calcium channel on
RT chromosome 17q and exclusion as candidate genes for malignant
RT hyperthermia susceptibility.";
RL HUM. MOL. GENET. 2:863-868(1993).
CC -!- FUNCTION: THIS PROTEIN IS THE BETA SUBUNIT OF THE DIHYDROPYRIDINE
CC (DHP) SENSITIVE CALCIUM CHANNEL. PLAYS A ROLE IN EXCITATION-
CC CONTRACTION COUPLING. THE SKELETAL MUSCLE DHP-SENSITIVE CA(2+)-
CC CHANNEL MAY FUNCTION ONLY AS A MULTIPLE SUBUNIT COMPLEX.
CC -!- SUBUNIT: THE SKELETAL MUSCLE L-TYPE CALCIUM CHANNEL IS COMPOSED OF
CC FIVE SUBUNITS: ALPHA-1, ALPHA-2, BETA, GAMMA AND DELTA.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ASSOCIATED WITH
CC THE CYTOPLASMIC ASPECT OF THE SARCOLEMMA.
CC -!- TISSUE SPECIFICITY: SKELETAL MUSCLE.
CC -!- ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1
CC SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M92301; G179802; -
DR EMBL: L06112; G187019; -
DR EMBL: U86960; G2155255; -
DR EMBL: U86952; G2155255; JOINED.
DR EMBL: U86953; G2155255; JOINED.
DR EMBL: U86954; G2155255; JOINED.
DR EMBL: U86955; G2155255; JOINED.
DR EMBL: U86956; G2155255; JOINED.
DR EMBL: U86957; G2155255; JOINED.
DR EMBL: U86958; G2155255; JOINED.
DR EMBL: U86959; G2155255; JOINED.
DR EMBL: Z21725; G38563; -
DR EMBL: Z21726; G38565; -
DR MIM: I14207; -
DR PROSITE: PS50002; SH3; 1.
DR PFAM: PF00018; SH3; 1.
DR PFAM: PF00774; Ca_channel_B; 1.
KW IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;
KW GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;
KW MULTIGENE FAMILY.
FT DOMAIN 100 161 SH3.
FT CARBOHYD 189 189 POTENTIAL.
FT CARBOHYD 470 470 POTENTIAL.
FT CONFLICT 28 29 QG -> R (IN REF. 2).
FT CONFLICT 183 183 MISSING (IN REF. 4).
FT CONFLICT 479 480 AA -> RR (IN REF. 1 AND 3).
SQ SEQUENCE 523 AA; 57863 MW; 32CBEE9D CRC32;

Query Match 79.48; Score 54; DB 1; Length 523;
Best Local Similarity 75.08; Pred. No. 3.03e-02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 44 STSDTTSNFV 55
QY 81 STSNDTTSAAFY 92

RESULT 3
ID CCB2_HUMAN STANDARD; PRT: 524 AA.
AC Q08289;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CALCIUM CHANNEL BETA-1M SUBUNIT
DE (BETA-1 ISOFORM C).
GN CACNB1 OR CACNLB1.
OS ORYCTOLAGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-SKELETAL MUSCLE;
RA RUTH P., ROEHRKASTEN A., BIEL M., BOSSE E., REGULLA S., MEYER H.F.,
RA FLOCKERZI V., HOFFMANN F.;
RT "Primary structure of the beta subunit of the DHP-sensitive calcium
channel from skeletal muscle.";
RL SCIENCE 245:1115-1118(1989).
CC -!- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROPYRIDINE (DHP)
SENSITIVE CALCIUM CHANNEL.
CC -!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:
ALPHA-1, ALPHA-2, BETA AND GAMMA.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ASSOCIATED WITH
THE CYTOPLASMIC ASPECT OF THE SARCOLEMMMA.
CC -!- ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1
SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE (BY
SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M25817; G164834; -
DR EMBL; M25514; G598370; -
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00774; Ca_channel_B; 1.
KW IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;
KW GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;
KW MULTIGENE FAMILY.
FT MOD_RES 71 ?1 BLOCKED.
FT DOMAIN 100 161 SH3.
FT CARBOHYD 189 189 POTENTIAL.
FT CARBOHYD 470 470 POTENTIAL.
FT CARBOHYD 499 499 POTENTIAL.
FT CONFLICT 304 304 L -> R (IN G598370).
SQ SEQUENCE 524 AA; 57825 MW; 8AAC8B67 CRC32;

Query Match 79.4%; Score 54; DB 1; Length 524;
Best Local Similarity 75.0%; Pred. No. 3.03e-02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 44 STSDTTSNFV 55
QY 81 STSNDTTSAAFY 92

RESULT 4

```

```

ID CCB2_HUMAN STANDARD; PRT: 567 AA.
AC Q08289;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CALCIUM CHANNEL BETA-2 SUBUNIT
DE (LAMBERT-EATON MYASTHENIC SYNDROME ANTIGEN B) (MYSB).
GN CACNB2 OR CACNLB2 OR MYSB.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-FETAL BRAIN;
RA ROSENFELD M.R., WONG E., DALMAU J., MANLEY G., POSNER J.B.,
RA SHER E., FURNEAUX H.M.;
RT "Cloning and characterization of a Lambert-Eaton myasthenic syndrome
antigen.";
RL ANN. NEUROL. 33:113-120(1993).
CC -!- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROPYRIDINE (DHP)
SENSITIVE CALCIUM CHANNEL (BY SIMILARITY).
CC -!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:
ALPHA-1, ALPHA-2, BETA AND GAMMA.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES.
CC -!- ALTERNATIVE PRODUCTS: THREE FORMS ARE PRODUCED BY ALTERNATIVE
SPLICING OF THE GENE.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S60415; G300417; -
DR PIR; A48895; A48895.
DR MIM; 600003; -
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00774; Ca_channel_B; 1.
KW IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;
KW GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;
KW ANTIGEN.
FT DOMAIN 114 175 SH3.
FT CARBOHYD 122 122 POTENTIAL.
FT CARBOHYD 203 203 POTENTIAL.
FT CARBOHYD 523 523 POTENTIAL.
FT VARSPLIC 234 268
FT VARSPLIC 224 268
FT VARSPLIC 224 268
FT VARSPLIC 224 268
SQ SEQUENCE 567 AA; 62087 MW; 4559C166 CRC32;

Query Match 79.4%; Score 54; DB 1; Length 567;
Best Local Similarity 75.0%; Pred. No. 3.03e-02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 58 STSDTTSNFV 69
QY 81 STSNDTTSAAFY 92

RESULT 5
ID CCB2_HUMAN STANDARD; PRT: 596 AA.
AC Q02641;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

```

DIHYDROXYRINDINE-SENSITIVE L-TYPE, CHANNEL BETA-1-B2 SUBUNIT
(BETA-1 ISOFORM A).
CACNB1 OR CACNLB1.
HOMO SAPIENS (HUMAN).
EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]
SEQUENCE FROM N.A.
TISSUE-HIPPOCAMPUS;
MEDLINE: 93054616;
POWERS P.A., LIU S., HOGAN K., GREGG R.G.;
"Skeletal muscle and brain isoforms of a beta-subunit of human
voltage-dependent calcium channels are encoded by a single gene.";
J. BIOL. CHEM. 267:22967-22972(1992).
[2]
SEQUENCE FROM N.A.
TISSUE-HEART;
MEDLINE: 93265672.
COLLIN T., WANG J., NARGEOT J., SCHWARTZ A.;
"Molecular cloning of three isoforms of the L-type voltage-dependent
calcium channel beta subunit from normal human heart.";
CIRC. RES. 72:1337-1344(1993).
[3]
-!- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROXYRINDINE (DRP)
SENSITIVE CALCIUM CHANNEL.
-!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:
ALPHA-1, ALPHA-2, BETA AND GAMMA.
-!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ASSOCIATED WITH
THE CYTOPLASMIC ASPECT OF THE SARCOLEMA.
-!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, HEART AND SPLEEN.
-!- ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1
SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE.
-!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
-!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; M92303; G179806; -;
EMBL; L06110; G187015; -;
PIR; C44461; C44461.
MIM; 114207; -;
PROSITE; PS50002; SH3; 1.
PFAM; PF00018; SH3; 1.
PFAM; PF00774; Ca_channel_B; 1.
IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;
GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;
MULTIGENE FAMILY.
FT DOMAIN 100 161 SH3.
FT CARBOHYD 189 189 POTENTIAL.
FT CARBOHYD 425 425 POTENTIAL.
FT CONFLICT 5 6 TS -> SG (IN REF. 2).
FT CONFLICT 29 29 G -> R (IN REF. 2).
FT CONFLICT 135 135 H -> D (IN REF. 2).
FT CONFLICT 175 176 KL -> TV (IN REF. 2).
FT CONFLICT 182 182 G -> S (IN REF. 2).
FT CONFLICT 217 217 T -> S (IN REF. 2).
FT CONFLICT 293 296 SNTR -> LQHT (IN REF. 2).
FT CONFLICT 344 344 I -> L (IN REF. 2).
FT CONFLICT 428 428 M -> I (IN REF. 2).
FT CONFLICT 434 435 AA -> RR (IN REF. 1).
FT CONFLICT 456 456 E -> D (IN REF. 2).
FT CONFLICT 465 465 M -> V (IN REF. 2).
FT CONFLICT 482 482 S -> W (IN REF. 2).
FT CONFLICT 492 492 R -> P (IN REF. 2).
FT CONFLICT 515 515 L -> P (IN REF. 2).
FT CONFLICT 532 532 L -> P (IN REF. 2).
FT CONFLICT 539 539 A -> GTP (IN REF. 2).
FT CONFLICT 548 548 MISSING (IN REF. 2).

FT CONFLICT 557 557 L -> M (IN REF. 2).
FT CONFLICT 571 571 WP -> CA (IN REF. 2).
FT CONFLICT 591 591 R -> Q (IN REF. 2).
SQ SEQUENCE 596 AA; 63638 MW; 0CC447D4 CRC32;
Query Match 79.4%; Score 54; DB 1; Length 596;
Best Local Similarity 75.0%; Pred. No. 3.03e-02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 44 STSDTTTSNSFV 55
|||:||||:|
Qy 81 STSDTTTSAAFV 92
RESULT 6
ID CCB_A_RAT STANDARD; PRT; 597 AA.
AC P34283;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE DIHYDROXYRINDINE-SENSITIVE L-TYPE, CHANNEL BETA-1-B2 SUBUNIT
(BETA-1 ISOFORM A).
GN CACNB1 OR CACNLB1.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RC SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 92038046.
RA PRAGNELL M., SAKAMOTO J., JAY S.D., CAMPBELL K.P.;
"Cloning and tissue-specific expression of the brain calcium channel
beta-subunit".
FEBS LETT. 291:253-258(1991).
CC -!- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROXYRINDINE (DRP)
SENSITIVE CALCIUM CHANNEL.
CC -!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:
ALPHA-1, ALPHA-2, BETA AND GAMMA.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ASSOCIATED WITH
THE CYTOPLASMIC ASPECT OF THE SARCOLEMA.
CC -!- ALTERNATIVE PRODUCTS: THREE, TISSUE-SPECIFIC, FORMS OF THE BETA-1
SUBUNIT ARE PRODUCED BY ALTERNATIVE SPLICING OF THE GENE (BY
SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; X61394; G55894; -;
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00774; Ca_channel_B; 1.
IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;
GLYCOPROTEIN; PHOSPHORYLATION; ALTERNATIVE SPLICING; SH3 DOMAIN;
MULTIGENE FAMILY.
FT DOMAIN 100 161 SH3.
FT CARBOHYD 189 189 POTENTIAL.
FT CARBOHYD 425 425 POTENTIAL.
SQ SEQUENCE 597 AA; 65679 MW; 23B08C47 CRC32;
Query Match 79.4%; Score 54; DB 1; Length 597;
Best Local Similarity 75.0%; Pred. No. 3.03e-02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 44 STSDTTTSNSFV 55
|||:||||:|
Qy 81 STSDTTTSAAFV 92

RESULT 7
ID CCB2_RABIT STANDARD; PRT; 632 AA.
AC P54288;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE DIHYDROPYRIDINE-SENSITIVE L-TYPE, CALCIUM CHANNEL BETA-2 SUBUNIT
DE (CAB2).
GN CACNB2 OR CACNLB2.
OS ORYCTOLAGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=HEART;
RC MEDLINE: 92192022.
RA HULLIN R., SINGER-LAHAT D., FREICHEL M., BIEL M., DASCAL N.,
RA HOFMANN F., FLOCKERZI V.;
RT "Calcium channel beta subunit heterogeneity: functional expression of
RT cloned cDNA from heart, aorta and brain.";
RL EMBO J. 11:885-890(1992).
CC -!- FUNCTION: THIS PROTEIN IS A SUBUNIT OF THE DIHYDROPYRIDINE (DHP)
CC SENSITIVE CALCIUM CHANNEL.
CC -!- SUBUNIT: THE L-TYPE CALCIUM CHANNEL IS COMPOSED OF FOUR SUBUNITS:
CC ALPHA-1, ALPHA-2, BETA AND GAMMA.
CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN HEART, AORTA AND
CC BRAIN.
CC -!- ALTERNATIVE PRODUCTS: THREE FORMS OF THE BETA-2 SUBUNIT ARE
CC PRODUCED BY ALTERNATIVE SPLICING OF THE GENE: CAB2A, CAB2B (SHOWN
CC HERE) AND CAB2C.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER CALCIUM CHANNEL BETA SUBUNITS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X64298; G1500; .
DR EMBL; X64297; G1498; .
DR EMBL; X64299; G1502; .
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00774; Ca_channel_B; 1.
KW IONIC CHANNEL; ION TRANSPORT; VOLTAGE-GATED CHANNEL; CALCIUM CHANNEL;
KW GLYCOPROTEIN; PHOSPHORYLATION; SH3 DOMAIN; MULTIGENE FAMILY;
KW ALTERNATIVE SPLICING.
FT DOMAIN 86 147 SH3.
FT CARBOHYD 94 94 POTENTIAL.
FT CARBOHYD 175 175 POTENTIAL.
FT VARSPIC 1 43
FT VARSPIC 197 234
FT VARSPIC 235 240 MISSING (IN CAB2C).
FT VARSPIC 235 240 MPFFKK -> KQKQKS (IN CAB2C).
SQ SEQUENCE 632 AA; 70943 MW; DF97DA5D CRC32;

Query Match 79.4%; Score 54; DB 1; Length 632;
Best Local Similarity 75.0%; Pred. No. 3.03e-02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 30 STSDTTSNSFV 41
QY 81 STSDTTSNSAFV 92

RESULT 8
ID LPFD_SALTY STANDARD; PRT; 355 AA.
AC P43663;

DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE LPFD PROTEIN PRECURSOR.
GN LPFD.
OS SALMONELLA TYPHIMURUM.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC SALMONELLA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 14028;
RX MEDLINE: 95238281.
RA BAUEMLER A.J., HEFFRON F.;
RT Identification and sequence analysis of lpfABCD, a putative
RT fimbrial operon of Salmonella typhimurium.*;
RL J. BACTERIOL. 177:2087-2097(1995).
CC -!- SIMILARITY: BELONGS TO THE FIMH / LPFD FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U18559; G829374; .
DR STYGENE; SG10560; LPFD.
DR PFAM; PF00419; Fimbrial; 1.
KW FIMBRIA; SIGNAL.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 355 LPFD PROTEIN.
SQ SEQUENCE 355 AA; 37714 MW; 20D529B5 CRC32;

Query Match 70.6%; Score 48; DB 1; Length 355;
Best Local Similarity 63.6%; Pred. No. 1.11e-00;
Matches 7; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 51 TSKNTTGATFV 61
QY 82 TSNDTTSAAFV 92

RESULT 9
ID EMRL_HUMAN STANDARD; PRT; 886 AA.
AC Q14246;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE CELL SURFACE GLYCOPROTEIN EMR1 PRECURSOR (EMR1 HORMONE RECEPTOR).
GN EMR1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95324926.
RA BAUD V., CHISSOE S.L., VIEGAS-PEQUIGNOT E., DIRIONG S., N'GUEN V.C.,
RA ROE B.A., LIPINSKI M.;
RT "EMR1, an unusual member in the family of hormone receptors with
RT seven transmembrane segments.";
RL GENOMICS 26:334-344(1995).
CC -!- FUNCTION: PROBABLY INVOLVED IN CELLULAR RESPONSE TO A HORMONE OR
CC AN INTERACTION WITH A PROTEIN LIGAND.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: WIDE EXPRESSION; INCREASED LEVELS IN
CC PERIPHERAL BLOOD MONONUCLEAR CELLS.
CC -!- PTM: N- AND O-GLYCOSYLATED; (POSSIBLE).
CC -!- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

WQERH

***** (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Sep 2 14:06:15 1999; MasPar time 1940.09 Seconds
Tabular output not generated. 1351.480 Million cell updates/sec

Title: >PCT-US99-13024-1
Description: (1-1119) from PCTUS9913024.seq
Perfect Score: 1119
N.A. Sequence: 1 atggaataattatgcccga.....aagaatttagaagaatttaa 1119
Comp: tacccttttaataacggct.....ttcttaattcttccaatt

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl-est58

1:em_est10 2:em_est11 3:em_est17 4:em_est18 5:em_est2
6:em_est9 7:em_gss1
genbank-est111

8:gb_est1 9:gb_est10 10:gb_est11 11:gb_est12 12:gb_est13
13:gb_est14 14:gb_est15 15:gb_est16 16:gb_est17
17:gb_est18 18:gb_est19 19:gb_est20 20:gb_est21
21:gb_est22 22:gb_est23 23:gb_est24 24:gb_est25
25:gb_est26 26:gb_est27 27:gb_est28 28:gb_est29
29:gb_est30 30:gb_est31 31:gb_est32 32:gb_est33
33:gb_est34 34:gb_est35 35:gb_est36 36:gb_est37
37:gb_est38 38:gb_est39 39:gb_gss3 40:gb_gss4 41:gb_gss5 42:gb_gss6

Statistics: Mean 11.199; Variance 2.112; scale 5.303

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.	
1	59	5.3	252	17	AA754459 97SN1787 Rice Immature	4.05e-63	
c	2	56	5.0	252	17	AA754459 97SN1787 Rice Immature	1.58e-57
3	54	4.8	247	17	AA754458 97SN1784 Rice Immature	7.72e-54	
c	4	51	4.6	247	17	AA754458 97SN1784 Rice Immature	2.26e-48
5	41	3.7	2275	20	AF034173 Human mRNA (T 1.70e-10)	7.60e-31	
c	6	28	2.5	2275	20	AF034173 Human mRNA (T 1.70e-10)	4.19e-09
7	27	2.4	507	28	AI506433 vm55e09.xl Stratiagene	4.19e-09	
c	8	27	2.4	595	15	AA596420 vm55e09.xl Stratiagene	4.19e-09
c	9	25	2.2	369	35	AA032210 zk19d04.s1 Soares.preg	2.00e-05
c	10	25	2.2	517	8	T58089 yb26a12.r1 Stratiagene	2.00e-05

C	11	23	2.1	251	24	AI205893
12	23	2.1	318	37	B96187	qf60a07.xl Soares_test
13	23	2.1	325	17	AA739545	T18ESTFB TAMU Arabidop
14	23	2.1	381	41	AA739545	310 PtiFG2 Pinus taeda
15	24	2.1	383	38	AA028987	RCI11-87A4.TJ RPC111
16	24	2.1	396	10	AA234431	CJT-HSP-2313A16.TR CIT
17	23	2.1	407	35	AA056859	z772a12.r1 Soares_Nhm
18	24	2.1	435	24	AI242393	SWMCA722SK Brugia mal
19	23	2.1	441	39	AQ113414	q336a09.xl NCI_CGAP_LY
20	23	2.1	447	19	T89700	CJT-HSP-2378L7.TR CIT-
21	23	2.1	453	40	AQ237532	Y999h08.s1 Soares_feta
22	23	2.1	454	41	AQ359215	HS_2019.B2.D08.MR CIT
23	23	2.1	475	39	AQ130893	HS_5029.B2.H08.SP6E RP
24	23	2.1	488	26	AI370893	HS_3036.A1.B09.MF CIT
25	23	2.1	502	9	AA151503	ta12405.xl Soares_tota
26	23	2.1	517	26	AI398575	z127b05.r1 Soares_preg
27	23	2.1	520	41	AQ334844	NCW01G1T7 Westergaards
28	23	2.1	527	17	AA651439	HS_5018.A1.F08.T7 RPCI
29	23	2.1	531	27	AI485437	31553 Lambda-PRL2 Arab
30	23	2.1	605	22	AU015321	EST24358 tomato ovary
31	24	2.1	606	41	AQ362692	AU015321 Mouse two-cel
32	23	2.1	637	33	W03126	nbxb0050006f CUGI Rice
33	24	2.1	654	39	AQ111972	z553c03.r1 Soares_feta
34	23	2.1	697	37	B16516	CJT-HSP-2377B1.TR CIT-
35	24	2.1	708	14	AA550589	342H10.TPB CIT978SKA1
36	24	2.1	764	14	AA550589	nbxb0060K14r CUGI Rice
37	24	2.1	799	41	AQ291506	1754m3 gmbPFB3.1, G.
38	23	2.1	997	39	AF010972	nbxb0025G01f CUGI Rice
39	22	2.0	331	17	AA704953	Homo sapiens chromosom
40	22	2.0	360	32	D75694	z195a01.s1 Soares_feta
41	22	2.0	372	15	C44854	CELK109AZF Yuji Kohara
42	22	2.0	449	23	AI175073	C44854 Yuji Kohara unp
43	22	2.0	609	25	AI294174	EST218591 Normalized r
44	22	2.0	722	37	B67669	LP07487.5prime LP Dros
45	22	2.0	812	41	AQ291855	T23N14TR TAMU Arabidop
						nbxb0040K19f CUGI Rice

ALIGNMENTS

RESULT 1 AA754459 252 bp mRNA EST 20-JAN-1998
LOCUS 97SN1787 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
DEFINITION CDNA clone 97SN1787, mRNA sequence.
ACCESSION AA754459
NID 92801165
VERSION AA754459.1 GI:2801165
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
REFERENCE 1 (bases 1 to 252)
AUTHORS Nahn,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
Lee,M.C. and Eun,M.Y.
TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
JOURNAL Unpublished (1998)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797457.
Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggi-do, Korea
Tel: 82 31 290 0301
Fax: 82 31 290 0307
Email: myeunsun20.asti.re.kr
Submitted by Baek Hie Nahn, Dept of Biological Science, Myongji
University, Yongin, Korea. 449-728 bnaahm@bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1..252
/organism="Oryza sativa"

```
/db_xref="taxon:4530"
/clone="97SN1787"
/clone_lib="Rice Immature Seed Lambda ZAPII CDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
/lab_xref="taxon:4530"
/clone="97SN1787"
/clone_lib="Rice Immature Seed Lambda ZAPII CDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
/lab_xref="taxon:4530"

BASE COUNT      5 a      21 c      12 g      35 t      179 others
ORIGIN

Query Match      5.3%; Score 59; DB 17; Length 252;
Best Local Similarity 11.8%; Pred. NO. 4.05e-63;
Matches 24; Conservative 109; Mismatches 66; Indels 4; Gaps 4;

Db 26 HGNVWCVASHGNYMNVHNCBTRGTHCDCKNVNWSMT-WGTVNVNBSGDMHYWBNTKVDVG 84
Qy 586 agtttgagcaattcatgtcgccagttgttggaactaatgacgaattgcgagattgccca 645
Db 85 TKVDVGNHTRCSRBRVTRAHYHDYTCBYYNNNDYHMMHBBMYBTGCMTCMWCWBH 144
Qy 646 acttcagctgataagataattctgattataatttaattctcgagattgtct-tta 704
Db 145 YNTKCTASGWHSTNYDVKSSTNTWGTBSYKSMHGYWCSBBVKYHTKYSTTRATRSYT 204
Qy 705 tcttaagctcagttatctgctgattgattctgcttcttggaactaca-aatatat 763
Db 205 CVRKYCVMMWTK-KVVKYHYVVB 226
Qy 764 ctatccgttttataatgcattct 786

RESULT 2
LOCUS AA754459 252 bp mRNA EST 20-JAN-1998
DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII CDNA Library Oryza sativa
CDNA clone 97SN1787, mRNA sequence.
ACCESSION AA754459
NID 92801165
VERSION AA754459.1 GI:2801165
KEYWORDS EST.
SOURCE Oryza sativa
ORGANISM Oryza sativa
REFERENCE 1 (bases 1 to 252)
AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
Lee,M.C. and Eun,M.Y.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
TITLE 1 (bases 1 to 252)
JOURNAL Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
COMMENT Unpublished (1998)
On Jan 14, 1998 this sequence version replaced gi:1797457.

Contact: Eun M.Y.
Department of Cytoogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
University, Yongin, Korea. 449-728 bhna@m Bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1. .252
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."

FEATURES
source
```

```
/db_xref="taxon:4530"
/clone="97SN1787"
/clone_lib="Rice Immature Seed Lambda ZAPII CDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
/lab_xref="taxon:4530"

BASE COUNT      5 a      21 c      12 g      35 t      179 others
ORIGIN

Query Match      5.0%; Score 56; DB 17; Length 252;
Best Local Similarity 12.1%; Pred. NO. 1.58e-57;
Matches 26; Conservative 109; Mismatches 76; Indels 3; Gaps 3;

Db 31 WVCVASHGNYMNVHNCBTRGTHCDCKNVNWSMTWGTVNVNBSGDMHYWBNTKVDVG 90
Cp 1014 aacacgaataccattgacgcaaatccagcaggttgcgaatgccattctctgatac 955
Db 91 NHTRCRSRBRVTRAHYHDYTCBYYNNNDYHMMHBB-BMYBBTGCMTCTMWCWBHNTKC 149
Cp 954 ttcagaaaagcttaattctctcatgacaccttgccagcttgctgtgcgtacgcta 895
Db 150 TA-SGWHSTNYDVKSSTNTWGTBSYKSMHGYWCSBBVKYHTKYSTTRATRSYTCVRK 208
Cp 894 tatcgtaaatcctactgcgtttcttggaacaaactacataatctcttaattcccatga 835
Db 209 YCVMMWTKVVK-KYHVVBSCBHTDSCKTKMM 241
Cp 834 cccagcttcccagtaaatcagctgtgttgaa 801

RESULT 3
LOCUS AA754458 247 bp mRNA EST 20-JAN-1998
DEFINITION 97SN1784 Rice Immature Seed Lambda ZAPII CDNA Library Oryza sativa
CDNA clone 97SN1784, mRNA sequence.
ACCESSION AA754458
NID 92801164
VERSION AA754458.1 GI:2801164
KEYWORDS EST.
SOURCE Oryza sativa
ORGANISM Oryza sativa
REFERENCE 1 (bases 1 to 247)
AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y.,
Lee,M.C. and Eun,M.Y.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
TITLE 1 (bases 1 to 247)
JOURNAL Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
COMMENT Unpublished (1998)
On Jan 14, 1998 this sequence version replaced gi:1797455.

Contact: Eun M.Y.
Department of Cytoogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggido, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.asti.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
University, Yongin, Korea. 449-728 bhna@m Bioserver.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1. .247
/organism="Oryza sativa"
/cultivar="Milyang23"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPII
vector at 5' end with EcoRI and 3' end with Xho I site."

FEATURES
source
```


This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:56400

Seq primer: -28m13 rev1 ET from Amersham

High quality sequence stop: 446.

Location/Qualifiers

FEATURES

1. 595
/organism="Mus musculus"
/note="Organ: blood; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. M30 CD4+ cells. Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGCGCAGG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3".
/db_xref="taxon:10090"
/clone="IMAGE:1002184"
/clone_lib="Stratagene mouse Tcell 937311"
/tissue_type="Tcell"
/dev_stage="M30 CD4+ cells"
/lab_host="SOLR (kanamycin resistant)"
BASE COUNT 158 a 135 c 119 g 182 t 1 others
ORIGIN

Query Match 2.4%; Score 27; DB 15; Length 595;
Best Local Similarity 78.7%; Pred. No. 4.19e-09;
Matches 37; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db 411 CTTGTGTTGCTGAATATACATTCAGATGCTTTGTGAACACGATA 457
||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Cp 811 cttgtgtgaagaataatccgttagcttagcattataaaacggata 765

RESULT

LOCUS AA032210 369 bp mRNA EST 21-AUG-1996
DEFINITION zk19d04.s1 Soares_pregnant_uterus_NbHPU Homo sapiens CDNA clone
IMAGE:470983 3', mRNA sequence.
ACCESSION AA032210
NID Q1502218
VERSION AA032210.1 GI:1502218
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 369)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On May 9, 1995 this sequence version replaced gi:802460.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40m13 fwd. from Amersham

High quality sequence stop: 277.

Location/Qualifiers

FEATURES

1. 369
/organism="Homo sapiens"
/note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo (dt) primer [5' AACTGGAAGATTTCGCGCGCCCTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library

went through one round of normalization. Library constructed by M. Fatima Bonaldo."

/db_xref="GDB:3756985"

/db_xref="taxon:9606"

/map="12p"

/clone="IMAGE:470983"

/clone_lib="Soares_pregnant_uterus_NbHPU"

/sex="female"

/dev_stage="adult"

/lab_host="DH10B"

BASE COUNT 134 a 69 c 58 g 107 t 1 others
ORIGIN

Query Match 2.2%; Score 25; DB 35; Length 369;

Best Local Similarity 65.8%; Pred. No. 2.00e-06;

Matches 50; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Db 246 TAAATAGCCATCAGATACATGGAATTTTAAGATGAGATTAACCTANTCCAGTATATT 305
||||| ||||||| ||| || ||||||| ||| ||| ||| |||

Cp 738 taaatcagcatcagcatataactgagctttaagataaagaacatctccaggaattaaatt 679

Db 306 ATACTCACAAATATA 321
||||| ||||||| ||| |||

Cp 678 ataacagataaattca 663

RESULT

LOCUS T58089 517 bp mRNA EST 08-FEB-1995
DEFINITION yb26a12.r1 Stratagene fetal spleen (#937205) Homo sapiens CDNA
clone IMAGE:72286 5' similar to similar to gb:M31468 RAS-LIKE
PROTEIN TC21 (HUMAN), mRNA sequence.
ACCESSION T58089
NID G659950
VERSION T58089.1 GI:659950
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 517)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
TITLE WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 887

High quality sequence stop: 312

Location/Qualifiers

1. 517
/organism="Homo sapiens"

/note="Organ: spleen; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:

Oligo dt. Pooled spleens. Average insert size: 1.0 kb;

Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGCGCAGG

3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3".

/db_xref="GDB:49351"

/db_xref="taxon:9606"

/map="17 cen-qter"

/clone="IMAGE:72286"

/clone_lib="Stratagene fetal spleen (#937205)"

```
/tissue_type="fetal spleen"
/dev_stage="fetal"
/lab_host="SOLR cells (kanamycin resistant)"
BASE COUNT      130 a      98 c      135 g      145 t      9 others
ORIGIN

Query Match      2.2%; Score 25; DB 8; Length 517;
Best Local Similarity 64.3%; Pred. No. 2.00e-06;
Matches 54; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Db 98 ATGAGCTGCCATCTGGACATTTTGGATACAGCTGGAGCAGAGCTTACAGCCATGC 157
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cp 427 atgatactgcaatcttgcaatagatagatcagatggaagcagacgttcgcggcttc 368
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 158 GNCACCATGATATGAGCGGCGAGGAG 181
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cp 367 ttaaccaagatactactacttcaggag 344
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
LOCUS      AI205893      251 bp      mRNA      EST      16-OCT-1998
DEFINITION qf60a07.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1754388
3', mRNA sequence.
ACCESSION  AI205893
NID         93764565
VERSION     AI205893.1 GI:3764565
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 251)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    On Jan 19, 1998 this sequence version replaced gi:2151858.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 245.
Location/Qualifiers
1..251
/organism="Homo sapiens"
/notes="Vector: pTT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="IMAGE:1754388"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
BASE COUNT      95 a      34 c      27 g      95 t
ORIGIN

Query Match      2.1%; Score 23; DB 24; Length 251;
Best Local Similarity 75.9%; Pred. No. 6.63e-04;
Matches 41; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

Db 125 TACAGTCTCAATATGAAÁ-CTAAAAAGTTTCGATAATCATTTAAGCATTTTCT 177
|| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Cp 543 taaactgtcaaatatacaatctaaatagtttgataatcatctgcttcttct 490
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
LOCUS      B96187      318 bp      DNA      GSS      30-MAR-1998
DEFINITION T18E5F8B TAMU Arabidopsis thaliana genomic clone T18E5, genomic
survey sequence.
ACCESSION  B96187
NID         g2998663
VERSION     B96187.1 GI:2998663
KEYWORDS    GSS.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
Eukaryota; Viridiplantae; Charophyta/Embryophyta group;
Embryophyta; Tracheophyta; seed plants; Magnoliophyta;
eudicotyledons; Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 318)
AUTHORS   Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and
Venter,J.C.
TITLE     A BAC End Sequence Database for Identifying Minimal Overlaps in
Arabidopsis Genomic Sequencing. Update 3
JOURNAL    Unpublished (1997)
COMMENT    Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@tigr.org
Seq primer: M13-21
Class: BAC ends
High quality sequence stop: 318.
Location/Qualifiers
1..318
/organism="Arabidopsis thaliana"
/strain="Columbia"
/notes="Vector: BelobACII; Site_1: HindIII; Site_2:
HindIII; Produced by Rod Wing"
/d_xref="taxon:3702"
/clone="T18E5"
/clone_lib="TAMU"
/sex="hermaphrodite"
BASE COUNT      93 a      42 c      51 g      131 t      1 others
ORIGIN

Query Match      2.1%; Score 23; DB 37; Length 318;
Best Local Similarity 92.6%; Pred. No. 6.63e-04;
Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 204 ATAATCGAATTTTCTGCTATAACTA 230
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 299 ataatcgaaattgttgctataatacta 325
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
LOCUS      AA739545      325 bp      mRNA      EST      15-JAN-1998
DEFINITION 310 PfIFG2 Pinus taeda cDNA clone 8507M 3', mRNA sequence.
ACCESSION  AA739545
NID         g2778096
VERSION     AA739545.1 GI:2778096
KEYWORDS    EST.
SOURCE      loblolly pine.
ORGANISM    Pinus taeda
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Coniferopsida; Coniferales;
```

http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html
 Seq primer: SP6
 Class: BAC ends

FEATURES
 source
 1. .381
 /organism="Homo sapiens"
 /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
 RPC111 Human Male BAC Library"
 /db_xref="GDB:7533027"
 /db_xref="taxon:9606"
 /clone="R-87A4"
 /clone_lib="RPC111"
 /sex="Male"
 /cell_type="Lymphocytes"
 121 a 56 c 140 t

BASE COUNT 121 a 56 c 140 t
 ORIGIN

Query Match 2.1%; Score 23; DB 41; Length 381;
 Best Local Similarity 79.5%; Pred. No. 6.63e-04;
 Matches 31; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Db 286 TATGTAAGATTTCTATTATCATATGCTGCTTTTATA 324
 ||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Qy 404 tattgtcaagattgcagtcatactatgctgttttata 442

RESULT 15
 LOCUS AQ028987 383 bp DNA GSS 29-JUN-1998
 DEFINITION CIT-HSP-2313A16.TR CIT-HSP Homo sapiens genomic clone 2313A16.
 genomic survey sequence.
 ACCESSION AQ028987
 NID 93269209
 VERSION AQ028987.1 GI:3269209
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 383)
 AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,
 Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
 Simon,M., and Venter,J.C.
 TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
 Building
 JOURNAL Unpublished (1997)
 COMMENT Other_GSSs: CIT-HSP-2313A16.TF
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org
 Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:
 http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html.
 Seq primer: M13 Reverse
 Class: BAC ends

FEATURES
 source
 1. .383
 /organism="Homo sapiens"
 /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
 HindIII"
 /db_xref="taxon:9606"
 /clone="2313A16"
 /clone_lib="CIT-HSP"
 /sex="Male"
 /cell_type="Sperm"
 127 a 74 c 50 g 132 t

BASE COUNT 127 a 74 c 50 g 132 t
 ORIGIN

Query Match 2.1%; Score 24; DB 38; Length 383;
 Best Local Similarity 71.4%; Pred. No. 3.82e-05;

Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 326 CTAATTCAGAGGAAATTCACATTTTACTATCCACTTTTAGTTACCATTCCTTA 381

Cp 355 ctacttcaggaggaaaaattaaactttccactagtagtaataatagcaacaattcgatta 300

Search completed: Thu Sep 2 14:51:10 1999

Job time : 2695 secs.

 M E R E L L (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
 Run on: Thu Sep 2 14:51:32 1999; MasPar time 249.17 Seconds
 Tabular output not generated. 962.282 Million cell updates/sec

Title: >PCT-US99-13024-1
 Description: (1-1119) from PCTUS9913024.seq
 Perfect Score: 1119
 N.A. Sequence: 1 atggaataattatgcccga.....agaatttagagaatttaa 1119
 Comp: taccttttaataaccggct.....ttctaaatctctcaatt

Scoring table: TABLE default
 Gap 6

Nmatch STD : Dbase 0: Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: n-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33
 34:part34 35:part35 36:part36 37:part37 38:part38
 39:part39 40:part40 41:part41 42:part42 43:part43
 44:part44 45:part45 46:part46 47:part47 48:part48
 49:part49 50:part50 51:part51 52:part52 53:part53
 54:part54 55:part55 56:part56 57:part57 58:part58
 59:part59 60:part60

Statistics: Mean 8.996; Variance 5.350; scale 1.682

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.	
1	1078	96.3	8855	24	T29033	Phage T4 tail fibre g	0.00e+00
2	39	3.5	91	9	Q51746	Oligonucleotide probe	8.60e-09
3	38	3.4	204	1	N81164	Base substituted E.co	3.23e-08
4	37	3.3	91	9	Q51746	Oligonucleotide probe	1.20e-07
5	37	3.3	204	1	N81164	Base substituted E.co	1.20e-07
6	36	3.2	91	46	V44650	Mammalian DNA replica	4.41e-07
7	31	2.8	114	12	Q70465	Generic DNA sequence	2.50e-04
8	30	2.7	91	46	V44650	Mammalian DNA replica	8.57e-04
9	30	2.7	114	12	Q70468	Generic DNA sequence	8.57e-04

10	29	2.6	114	12	Q70469	Generic DNA sequence	2.89e-03
11	29	2.6	114	12	Q70466	Generic DNA sequence	2.89e-03
12	29	2.6	114	12	Q70467	Generic DNA sequence	2.89e-03
13	29	2.6	114	12	Q70469	Generic DNA sequence	2.89e-03
14	28	2.5	114	12	Q70466	Generic DNA sequence	9.59e-03
15	27	2.4	39	7	Q51787	Mixed oligonucleotide	3.13e-02
16	27	2.4	114	12	Q70467	Generic DNA sequence	3.13e-02
17	27	2.4	114	12	Q70465	Generic DNA sequence	3.13e-02
18	27	2.4	501	3	N50024	Sequence encoding new	3.13e-02
19	27	2.4	501	3	N50026	Sequence encoding new	3.13e-02
20	26	2.3	39	7	Q51787	Mixed oligonucleotide	1.00e-01
21	26	2.3	114	12	Q70468	Generic DNA sequence	1.00e-01
22	25	2.2	498	3	N50034	Sequence encoding new	3.15e-01
23	25	2.2	501	3	N50030	Sequence encoding new	3.15e-01
24	25	2.2	501	3	N50023	Sequence encoding new	3.15e-01
25	25	2.2	501	3	N50032	Sequence encoding new	3.15e-01
26	25	2.2	501	3	N50029	Sequence encoding new	3.15e-01
27	25	2.2	501	3	N50027	Sequence encoding new	3.15e-01
28	25	2.2	501	3	N50031	Sequence encoding new	3.15e-01
29	25	2.2	501	3	N50025	Sequence encoding new	3.15e-01
30	25	2.2	501	3	N50033	Sequence encoding new	3.15e-01
31	25	2.2	501	3	N50028	Sequence encoding new	3.15e-01
32	25	2.2	501	3	N50025	Sequence encoding new	3.15e-01
33	24	2.1	68	34	T73397	Oligonucleotide tag c	9.69e-01
34	24	2.1	68	32	T63255	Messenger RNA primer	9.69e-01
35	24	2.1	114	12	Q70472	Generic DNA sequence	9.69e-01
36	24	2.1	114	12	Q70470	Generic DNA sequence	9.69e-01
37	24	2.1	498	3	N50034	Sequence encoding new	9.69e-01
38	24	2.1	501	3	N50028	Sequence encoding new	9.69e-01
39	24	2.1	501	3	N50029	Sequence encoding new	9.69e-01
40	24	2.1	501	3	N50024	Sequence encoding new	9.69e-01
41	24	2.1	501	3	N50031	Sequence encoding new	9.69e-01
42	24	2.1	501	3	N50023	Sequence encoding new	9.69e-01
43	24	2.1	501	3	N50026	Sequence encoding new	9.69e-01
44	24	2.1	501	3	N50032	Sequence encoding new	9.69e-01
45	24	2.1	855	60	X13639	Enterococcus faecalis	9.69e-01

ALIGNMENTS

RESULT 1
 ID T29053 standard; DNA; 8855 BP.
 AC T29053;
 DT 07-JAN-1997 (first entry)
 DE Phage T4 tail fibre genes.
 KW Phage T4; tail fibre protein; nanotechnology; nano-structure;
 KW filter; molecular sieve; ss.
 OS Bacteriophage T4.
 FH Key Location/Qualifiers
 FT cds 16..3885
 FT /*tag= a
 FT /label= Gene_34
 FT /product= gp34 protein
 FT 3894..4091
 FT /*tag= b
 FT /label= ORFX
 FT /product= unidentified product
 FT 4127..5014
 FT /*tag= c
 FT /label= Gene_35
 FT /product= gp35 protein
 FT 5077..5742
 FT /*tag= d
 FT /label= Gene_36
 FT /product= gp36 protein
 FT 5751..8831
 FT /*tag= e
 FT /label= Gene_37
 FT /product= gp37 protein
 PN WP09611947.AL.
 PD 25-APR-1996.
 PF 13-OCT-1995; UI3023.
 PR 13-OCT-1994; US-322760.


```

OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
FT /tag= a
FT /function= multiple cloning site
FT primer_bind 187..204
FT /tag= b
PN EP-285123-A.
PD 05-MAY-1988.
PR 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prep of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; P; English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 3.4%; Score 38; DB 1; Length 204;
Best Local Similarity 11.2%; Pred. No. 3.23e-08;
Matches 11; Conservative 50; Mismatches 37; Indels 0; Gaps 0;

Db 88 rtthhyrmbvdyrnsdaawccyrrsvkdyccynachdhdybbvbyvnhn 147
Cp 759 atttgtagttccagagcaagtaatacagcatcataactagcgtttaagataag 700
Db 148 hnnccnccbnhchvbnhnnrwnayvrhddrdv 185
Cp 699 aacatctccagggaattaaattataatcagataattac 662

RESULT 4
ID Q51746 standard; cDNA; 91 BP.
AC Q51746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PR 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT ) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
DR WPI: 93-378844/48.
PT New oligo:nucleotide probes specific for Mycobacterla - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 3.3%; Score 37; DB 9; Length 91;
Best Local Similarity 10.5%; Pred. No. 1.20e-07;

Matches 6; Conservative 41; Mismatches 10; Indels 0; Gaps 0;

Db 2 gctccgcgsgsvhsvyvvshhshhsvhvhvsvvvvhhvvhvvhvhyv 58
QY 170 gcgagggccctcccttaagaggttttgacctccaccgagcattagttagta 226

RESULT 5
ID N81164 standard; DNA; 204 BP.
AC N81164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
FT /tag= a
FT /function= multiple cloning site
FT primer_bind 187..204
FT /tag= b
PN EP-285123-A.
PD 05-MAY-1988.
PR 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prep of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; P; English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 3.3%; Score 37; DB 1; Length 204;
Best Local Similarity 9.7%; Pred. No. 1.20e-07;
Matches 11; Conservative 57; Mismatches 45; Indels 0; Gaps 0;

Db 73 aayvcchvccgymttthhyrmbvdyrnsdaawccyrrsvkdyccynachh 132
QY 226 aegcatatgccttcttcgacctcaaatgatactacatcagctgtttgttagtcatg 285
Db 133 ddhyvbbvynvnhnncnccbnhchvbnhnnrwnayvrhddrdv 185
QY 286 aattcttcagcaataatgaattgtgtctattataatcagtagtgaaagggttaa 338

RESULT 6
ID V44650 standard; DNA; 91 BP.
AC V44650;
DT 06-OCT-1998 (first entry)
DE Mammalian DNA replication origin consensus sequence, uniorsconsensus.
KW DNA replication origin; human; mammal; alphaconsensus; uniorsconsensus;
KW anti-gene; DNA replication inhibitor; shuttle vector construct creation;
KW gene therapy; ss.
OS Mammalia.
PN WO9827200-A2.
PD 25-JUN-1998.
PR 12-DEC-1997; CA0972.
PR 21-MAY-1997; US-047322.
PR 16-DEC-1996; US-033374.
PA (UYMC-) UNIV MCGILL.

```

W O R L D
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Thu Sep 2 12:58:29 1999; MasPar time 2053.96 Seconds
Tabular output not generated.
1510.140 Million cell updates/sec

Title: >PCT-US99-13024-1
Description: (1-1119) from PCTUS9913024.seq
Perfect Score: 1119
N.A. Sequence: 1 atggagaaatttatgcccga.....aagaatttagagaatttaa 1119
Comp: taccttttaataaccgct.....ttcttaatactcttcaatt

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb158

1:em_ba1 2:em_ba2 3:em_fun 4:em_htg 5:em_hum1 6:em_hum2
7:em_in 8:em_om 9:em_of 10:em_ov 11:em_pat 12:em_ph
13:em_pl 14:em_ro 15:em_sts 16:em_v1

Database: genbank11
17:gb_ba1 18:gb_ba2 19:gb_htg1 20:gb_htg2 21:gb_in1
22:gb_in2 23:gb_om 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pl1
28:gb_pl2 29:gb_pr1 30:gb_pr2 31:gb_pr3 32:gb_ro
33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_v1

Statistics: Mean 10.969; Variance 5.248; scale 2.090

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	176	15.7	4418 26	PT4G36G37	Bacteriophage T4 genes
2	176	15.7	4418 26	MY74TL	Phage T4 tail fiber ge
3	156	13.9	902 26	MYOX2F36	Bacteriophage OX2 tail
4	152	13.6	911 26	MY2F36	Bacteriophage T2 tail
5	152	13.6	944 26	MYK3F36	Bacteriophage K3 tail
6	50	4.5	7218 25	I66494	Sequence 14 from paten
7	44	3.9	7218 25	I66494	Sequence 14 from paten
8	37	3.3	10772 21	AF012089	Drosophila melanogaste
9	37	3.3	10772 21	AF012089	Drosophila melanogaste
10	34	3.0	354 21	OFU89259	Oxytricha fallax 57kd
11	32	2.9	215 25	I28278	Sequence 5 from patent
12	33	2.9	370 21	OFU89262	Oxytricha fallax 57kd
13	33	2.9	965 25	AR024229	Sequence 22 from patent

C	14	33	2.9	216021	31	HUAC004787	Homo sapiens Chromosom	4.81e-05
C	15	31	2.8	216021	31	HUAC004787	Homo sapiens Chromosom	8.22e-04
C	16	30	2.7	74371	31	AC005369	Homo sapiens Chromosom	3.29e-03
C	17	29	2.6	1056	23	MYU87256	Mustela vison GT dinuc	1.29e-02
C	18	29	2.6	74371	31	AC005369	Homo sapiens Chromosom	1.29e-02
C	19	26	2.3	354	21	OFU89259	Oxytricha fallax 57kd	6.61e-01
C	20	26	2.3	641	23	MYU87253	Mustela vison GT dinuc	6.61e-01
C	21	26	2.3	965	25	AR024229	Sequence 22 from paten	6.61e-01
C	22	26	2.3	7207	22	CELT2584	Caenorhabditis elegans	6.61e-01
C	23	26	2.3	79073	27	AB008271	Arabidopsis thaliana g	6.61e-01
C	24	26	2.3	140306	19	HS1068F16	Human DNA sequence ***	6.61e-01
C	25	26	2.3	158166	19	AC005813	Drosophila melanogaste	6.61e-01
C	26	25	2.2	23636	21	CELF47C10	Caenorhabditis elegans	2.32e+00
C	27	25	2.2	86004	29	HS1J723	Human DNA sequence fro	2.32e+00
C	28	24	2.1	215	25	I28278	Sequence 5 from patent	7.88e+00
C	29	24	2.1	886	31	AF084462	Homo sapiens GIP-bindl	7.88e+00
C	30	24	2.1	942	30	HSU71203	Human rit mRNA, comple	7.88e+00
C	31	24	2.1	1037	30	HSU78165	Human small G protein	7.88e+00
C	32	24	2.1	1056	23	MYU87256	Mustela vison GT dinuc	7.88e+00
C	33	24	2.1	1112	30	HSRITGENE	H. sapiens mRNA for RIT	7.88e+00
C	34	24	2.1	1189	17	CLOORFCD	C.difficile DNA fragme	7.88e+00
C	35	24	2.1	5105	17	CDTOXUTX	C.difficile toxB and t	7.88e+00
C	36	24	2.1	9770	17	CDTOXA	Clostridium difficile	7.88e+00
C	37	24	2.1	10068	32	D89593	unannotatable data.	7.88e+00
C	38	24	2.1	26039	17	CD26039	C.difficile cdu2, cdu1	7.88e+00
C	39	24	2.1	31812	27	SPUNK4	S.pombe chromosome I c	7.88e+00
C	40	24	2.1	37685	30	HSAC000371	Human cosmid g1980a181	7.88e+00
C	41	24	2.1	43309	30	AC002498	Human Cosmid g5129g124	7.88e+00
C	42	24	2.1	43407	31	AC005584	Homo sapiens clone UWG	7.88e+00
C	43	24	2.1	115905	29	HS57E3	Homo sapiens DNA seque	7.88e+00
C	44	24	2.1	137557	31	AC005909	Homo sapiens chromsom	7.88e+00
C	45	23	2.1	156378	20	AC006453	Homo sapiens clone DJ0	2.58e+01

ALIGNMENTS

RESULT	1	PT4G36G37	4418 bp	DNA	PHG	15-SEP-1990
LOCUS		Bacteriophage T4 genes 36 and 37 (complete); 35 and 38 (partial)				
DEFINITION		Encoding tail fiber proteins.				
ACCESSION		J02509 J02508				
NID		G215884				
VERSION		J02509.1 GI:215884				
KEYWORDS		tail fiber protein.				
SOURCE		Bacteriophage T4				
ORGANISM		Bacteriophage T4				
REFERENCE		1 (bases 1 to 4418)				
AUTHORS		Oliver,D.B. and Crowther,R.A.				
TITLE		DNA sequence of the tail fibre genes 36 and 37 of bacteriophage T4				
JOURNAL		J. Mol. Biol. 153, 545-568 (1981)				
MEDLINE		82170495				
REFERENCE		2 (bases 180 to 235)				
AUTHORS		Christensen,A.C. and Young,E.T.				
TITLE		T4 late transcripts are initiated near a conserved DNA sequence				
JOURNAL		Nature 299, 369-371 (1982)				
MEDLINE		82272427				
COMMENT		Tail fiber genes gp36 and gp37 code for the distal half of the tail fiber of phage T4. The thin tip of the distal half fiber interacts with the bacterial lipopolysaccharide receptor and specifies the host range of the phage. The r at position 4235 in [1] should be a g in order to encode glutamic acid at that codon as stated in [1]. Location/Qualifiers 1. .4418 /organism="Bacteriophage T4" /db_xref="taxon:10665" /note="tail fiber protein 35" /codon_start=2 /transl_table=11 /protein_id="AAA32512.1"				
FEATURES		source				
		CDS				

	/db_xref="PID:g215885"	
	/db_xref="GI:215885"	
	/translation="SFSEVSRNGGISKPAEFGVGIRVNICESASPPDIMVLPTQAS	
CDS	SKTGVGGEFREV"	
	241..906	
	/note="tail fiber protein 36"	
	/codon_start=1	
	/transl_table=11	
	/protein_id="AAA32513.1"	
	/db_xref="PID:g215886"	
	/db_xref="GI:215886"	
	/translation="MADLKVGSTTGGSVIWHQGNFLPNDAGDDVLVKSKFIYSEYKPK	
	QAANDFVSKANGGTAYASKYTFNAGIQVPYAPNIMSPCGIYGNGGGAFFDKANIDIV	
	SWTGVEKFSFGSTRVRVINTNRGDNITKVVSAAQOVRSAGAAPAIANDLTKRKYD	
	DGAINTVTANANSRVLRSGDTMTGLTNPNFFSQNPASPQHVPREFDIQIKDSVDQF	
	GYY"	
CDS	915..3995	
	/note="tail fiber protein 37"	
	/codon_start=1	
	/transl_table=11	
	/protein_id="AAA32514.1"	
	/db_xref="PID:g215887"	
	/db_xref="GI:215887"	
	/translation="MATLKOIFKRSKIAGTRPAASVLAEGELAINLKORTIFTKKDDS	
	GNIIDLGFAKGQGDNVINGLLRLNGDYVQTGMTVNGPIGSTDGVTKGRIFRSTQG	
	RSFYARINETSIAHLFPENADGETGVYIARPOTTTDGEIRLVRVROGTGSTANSEYF	
	SFTNGEQFANRLIASDLVTKRIAYDVITHDAKFQCYDSDHSLVNVYPGTGETGV	
	NLYURKYAARGSGGT IYHEIVTAQTLGADVESWNSGDDTPVEKLIGRDDGRMIIRNSIAL	
	GTETTFNRSSDYGVMGDKYLVDGTVTGLSYKKTGYFDLVGGCYSVASITPDPSFR	
	TPBKGIJFGSREDOGATWIPGTNAALLSVQTQADNNAGDGQTHICYNAGKMNHTYFR	
	GTQMINSVCQGHREINPLKLVLTGNNVOFYADGTISSIQPIKLDEILFLTKNNTA	
	GLKAFPAQVDQVTDITOMWGDTREGQNKNYIIKAWGSFNFNATGDRSRETVEFQSDS	
	GVYFYAHRAKPTGDETIGREAQFDAGDYAKGIANGNFRVVGSSALAGNVMTSNGLF	
	VOCSSGITQVYKIGITANALRIWNAEYGAIFRRSENFYIIPTNQESGDDIHSHSLR	
	PVRIGLNDGMVGLGRDSFYVDONNALTTINSNSRIANFRMOLGOSAYIDAECTDAVR	
	PACAGSFASONEDVRAPFYMNIIRDRTDASNYPILKQRYVQNGCYSLSGTLNGNFR	
	VHYHGDDNGSTGPQTADGWEIFKNGDFISPRDLIAGKVRDRTNITGGSGGNFANL	
	NSTIESLKTDIMSSYPIGAPIPWPSDFVAGFALMEQGTQFKTSYAPKLAYVPSPVIP	
	DMRGQTIYIKKPSGRAVLAEADQVAKAHSASASSTD LGTKTSSFDYDGT KGTGNTSGTG	
	HTHSGSDYSTNGSEHSYIEAWNGTGVGKNKSSYAISYRAGGNTNAAGNISHTSF	
	GTSAGDSHSHSVSIGAHTHTVAIGSHGHTITVNSTGNTENTYKNIAFNIVRLA"	
CDS	4022..>4418	
	/note="tail fiber protein 38"	
	/codon_start=1	
	/transl_table=11	
	/protein_id="AAA32515.1"	
	/db_xref="PID:g553025"	
	/db_xref="GI:553025"	
	/translation="MKIYHYFTKEFYKEENYKPVKGLGLPAHSTIKKPLEPKGYA	
	VFEDETQMWIEEDHRGRAMTFNKEEIFISDGSFVGITTEDEGFEDIWDDGKWKE	
	DET YKRVLRNKRRIEELYEFQVNLNNIEA"	
BASE COUNT	1347 a 799 c 976 g 1296 t	
ORIGIN	1 bp upstream of HindIII site; 153.4 kb on genomic map.	
Query Match	15.7%; Score 176; DB 26; Length 4418;	
Best Local Similarity	99.4%; Pred. No. 4.71e-119;	
Matches	177; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Db	1 AACGTTTTCTCAAGATAACGCGCATTCACAAACCCTGCTGAATTGGCGTCAA 60	
Qy		
Qy	942 aagcttttcgaagtccaagaataggcgacattcgcgaacctgctaattggcgctcaa 1001	
Db	61 TGGTATTTCGTGTTAAATATATCTGCGAATCCGGCTTACCTCCCGATATAATGGTACTTCC 120	
Qy		
Qy	1002 tggtatttcglttaattatctcgcaatccgcttcacccccggatataataaggacttcc 1061	
Db	121 TAGCCAAGCATCGTCTAAAACTGGTAAAGTGTTTGGGCCAAGAATTTTAGAGAAGTTTAA 178	
Qy		
Qy	1062 tacgcagcatctctaaactggtaaaagtgttgggcagaagatttagagaagttaa 1119	
RESULT	2 MYT4TL PHG 07-JUL-1995	
LOCUS		

```

/db_xref="GI:15375"
/db_xref="SWISS-PROT:P03739"
/translation="MKIHYFDTKFYKEENYKPVKGLGLPAHSTIKKPLSPKEGYA
VFDEDTQWIEEDHURKRAWTFNKXIFISDTSIGSPVGTIDEPCEFDIWDGKWE"
DEFYKVLRLNRKRIEELYKEFVLNNMIEA"

BASE COUNT   1348 a   797 c   975 g   1296 t      2 others
ORIGIN

Query Match      15.7%   Score 176;   DB 26;   Length 4418;
Best Local Similarity 99.4%;   Pred. No. 4.71e-119;
Matches 177;   Conservative 0;   Mismatches 1;   Indels 0;   Gaps 0;

Db 1  AAGCTTTTCTGAAGTATCAGAAATGCGGGGATTTCGAAACCTGCTGCTGAATTTGGCGTCAA 60
|||||
Qy 942  aagcttttctgaagtacaagaatgagcgccatttcgaacctgctgaattggcgtaa 1001
|||||

Db 61  TGGTATTCTGTTTAATTATATCTCGGAATCGCTTCACCTCCGGATATATGCTACTTCC 120
|||||
Qy 1002  tggatttcgtagtaattatcttcgcgaatccgctccaccgccgataatggtactcc 1061
|||||

Db 121  TAGCGAAGCATCGTCTAAAACTGGTAAAGTGTTTGGCGAAGAATTTAGAGAAGTTTAA 178
|||||
Qy 1062  tagcgaagcatgctctaaactggtaaagtgtttggcgcaagaatttagagaagttaa 1119
|||||

RESULT 3
LOCUS      MYOX2F36              902 bp      DNA              PHG              24-JUL-1995
DEFINITION Bacteriophage Ox2 tail fiber gene 36.
ACCESSION  X01753
NID        G15122
VERSION    X01753.1  GI:15122
KEYWORDS   fiber protein; inverted repeat.
SOURCE     Bacteriophage Ox2.
ORGANISM   Bacteriophage Ox2.
            Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
            T4-like phages.
REFERENCE  1  (bases 1 to 902)
AUTHORS   Riede,I., Drexler,K. and Eschbach,M.L.
TITLE     The nucleotide sequences of the tail fiber gene 36 of bacteriophage
          T2 and of genes 36 of the T-even type Escherichia coli phages K3
          and Ox2
JOURNAL   Nucleic Acids Res. 13 (2), 605-616 (1985)
MEDLINE   85215500
FEATURES
     source              Location/Qualifiers
     1..902
     /organism="Bacteriophage Ox2"
     /db_xref="taxon:10891"
     1..178
     /gene="tail fiber gene 35"
     <1..178
     /gene="tail fiber gene 35"
     /codon_start=2
     /transl_table=11
     /protein_id="CAA25890.1"
     /db_xref="PID:e7867"
     /db_xref="PID:g1364176"
     /db_xref="GI:1364176"
     /db_xref="SPTREMBL:Q38400"
     /translation="SPSEVSRNGGISKPAEFVGNGIRVNYVCESASPPDIMVLPTQAS
     SATGKVFQGFREV"
     184..201
     /note="put. transcription terminator"
     184..190
     /note="repeat A"
     /rpt_type=INVERTED
     195..201
     /note="repeat A'"
     /rpt_type=INVERTED
     210..222
     /note="putative jukebox for late transcription Initiation
     (g36,g37,g38)"
     229..232
     /note="Shine Dalgarno sequence"

```

```
/gene="tail fiber gene 35"
<1..178
/gene="tail fiber gene 35"
/codon_start=2
/transl_table=11
/protein_id="CAA25896.1"
/db_xref="PID:g15190"
/db_xref="GI:15190"
/db_xref="SPTREMBL:Q38425"
/translation="SFSEVRNGSISKPAEFGVNGIRVNVVCESASPPDIIMVLPTQAS
SKTGKVFQGEFREV"
misc_signal
184..201
/note="put. transcription terminator"
repeat_unit
184..190
/note="imp. repeat A"
/rpt_type=INVERTED
repeat_unit
195..201
/note="imp. repeat A"
/rpt_type=INVERTED
promoter
210..222
/note="put."
RBS
229..232
/note="rRNA binding site"
gene
241..900
/gene="tail fiber gene 36 (aa 1-219)"
241..900
/gene="tail fiber gene 36 (aa 1-219)"
/codon_start=1
/transl_table=11
/protein_id="CAA25897.1"
/db_xref="PID:g15191"
/db_xref="GI:15191"
/db_xref="SPTREMBL:Q38426"
/translation="MADLVKVGSTGGSVIWHQGNFPLNPGADDDVLYKSKFIYSEYNKP
QAADNLVSKANGTGLNVRVGNRIQVPGAGNLTGMFVGCDGATDNVLDIISW
YGLGKSSOGTGPRTIVNVRGEISARGNINSORQVRAEAAAPTANDLTNRKDIYDG
AINTVTANANSRLVSGDTNLTAPNFFSQNPASQPSHVPRFDQIVIKDSVQDFGY
Y"
old_sequence
720
/gene="tail fiber gene 36 (aa 1-219)"
/citation={1}
/replace="t"
909..911
/gene="tail fiber gene 37"
909..911
/gene="tail fiber gene 37"
/codon_start=1
/transl_table=11
/protein_id="CAA25898.1"
/db_xref="PID:e7916"
/db_xref="PID:g4379275"
/db_xref="GI:4379275"
/translation="M"
BASE COUNT 269 a 171 c 200 g 271 t
ORIGIN

Query Watch 13.6% Score 152; DB 26; Length 911;
Best Local Similarity 92.7% Pred. No. 1.17e-98;
Matches 165; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Db 1 AACCTTTCTGAGGTATCAAGAAATGGTAGTATTTCGAACCCGCTGAATTTGGTGTCAA 60
QY 942 aagctttctgaagtatcaagaataagcgccatttcgaacacctgctgaattggcgtcaa 1001

Db 61 TGGTATTCGAGTTAAATATGCTGTGATCTGCTTCACCTCCGGATATATGCTACTTC 120
QY 1002 tggattctgtaataatatatctgcgaatccgcttcaccccggaataatggacttcc 1061

Db 121 TACACAAGCATCGCTAAACATGGTAAAGTGTGTTGGCGAAGATTTAGAGAAGTATAA 178
QY 1062 tacgcaagcatcgctctaaactggttaagtggtttggcggaagaatttagagaagtttaa 1119
```

```
5
LOCUS MYK3F36 944 bp DNA PHG 24-JUL-1995
DEFINITION Bacteriophage K3 tail fiber gene 36.
ACCESSION X01754
NID 915108
VERSION X01754.1 GI:15108
KEYWORDS fiber protein; inverted repeat.
SOURCE Bacteriophage K3.
ORGANISM Bacteriophage K3.
REFERENCE 1 (bases 1 to 944)
AUTHORS Riede,I., Drexler,K. and Eschbach,M.L.
TITLE The nucleotide sequences of the tail fiber gene 36 of bacteriophage
T2 and of genes 36 of the T-even type Escherichia coli phages K3
and OX2
JOURNAL Nucleic Acids Res. 13 (2), 605-616 (1985)
MEDLINE 85215500
FEATURES
source
Location/Qualifiers
1..944
/organism="Bacteriophage K3"
/db_xref="taxon:10674"
1..178
/gene="tail fiber gene 35"
<1..178
/gene="tail fiber gene 35"
/codon_start=2
/transl_table=11
/protein_id="CAA25893.1"
/db_xref="PID:e7855"
/db_xref="PID:g1364175"
/db_xref="GI:1364175"
/db_xref="SPTREMBL:Q38393"
/translation="SFSEVRNGSISKPAEFGINGIRVNVVCESASPPDIIMVLPTQAS
SKTGKVFQGEFREV"
misc_signal
184..201
/note="put. transcription terminator"
repeat_unit
184..190
/note="repeat A"
/rpt_type=INVERTED
repeat_unit
195..201
/note="repeat A"
/rpt_type=INVERTED
promoter
210..222
/note="putative jukebox for late transcription initiation
(936,937,938)"
RBS
229..232
/note="Shine Dalgarno sequence"
241..933
/gene="tail fiber gene 36 (aa 1-230)"
241..933
/gene="tail fiber gene 36 (aa 1-230)"
/codon_start=1
/transl_table=11
/protein_id="CAA25894.1"
/db_xref="PID:g15109"
/db_xref="GI:15109"
/db_xref="SWISS-PROT:P07626"
/translation="MADLVKVGSTGGSVIWHQGNFPLNPGADDDVLYKSKFIYSEYNKP
QATDNDFVSKANGTKSQGFQKEVREGVKISATISGSGDLNGLYSGNGDGSAREKA
NMDLRSMYIGIWNITCTGDSGTHNSGLMPGVVHVETCWRLLFOTDNVISNAGSPTGP
AHLTRKYDVGAINTVTANANSRLVSGDTNLTAPNFFSQNPASQPSHVPRFDQI
VIKDSVQDFGY"
942..944
/gene="tail fiber gene 37"
942..944
/gene="tail fiber gene 37"
/codon_start=1
/transl_table=11
/protein_id="CAA25895.1"
/db_xref="PID:e7857"
/db_xref="PID:g4379273"
/db_xref="GI:4379273"
/translation="M"
```

[illegible]

exon	2829 a	2357 c	2282 g	3046 t	158 others
exon	2829 a	2357 c	2282 g	3046 t	158 others
mRNA	2829 a	2357 c	2282 g	3046 t	158 others
CDS	2829 a	2357 c	2282 g	3046 t	158 others
Query Match	3.3%	Score 37;	DB 21;	Length 10772;	
Best Local Similarity	19.2%	Pred. No. 1.32e-07;			
Matches	29;	Conservative 70;	Mismatches 50;	Indels 2;	Gaps 2;
Db	1589	AAAKRWYIAWHTTWKWKMTTWKWKWKYRTTWKWKMTWSTTTTTSAMWYTWSTW	1748		
Cp	799	aaatatccggttagatgcattataaagcagtagatatatt-tgtagttccagagca	741		
Db	1749	TKYWWAYAWKMKWTRTWARMAWASARWKWKTSAAAYASAWRKWKWKWYARWKKTMMW	1808		
Cp	740	agtaaaaca-gcacagcagataactgactgactgaataaagacatctccaggaattaa	682		
Db	1809	AAWKRWKAAWMTWRWYMTTWTAAWRAAGCT	1839		
Cp	681	attataatcagataaattacttatcagcagt	651		
RESULT	10				
LOCUS	OFU89259	354 bp	DNA	INV	14-MAR-1997
DEFINITION	Oxytricha fallax 57kd zinc finger/protein chlmra gene, partial cds.				
ACCESSION	U89259				
NID	g1881675				
VERSION	089259.1	GI:1881675			
KEYWORDS					
SOURCE					
ORGANISM	Oxytricha fallax.				
REFERENCE	Eukaryotae; mitochondrial eukaryotes; Alveolata; Ciliophora; hypotriches; Stichotrichida; Oxytricha.				
AUTHORS	1 (bases 1 to 354)				
TITLE	Doak,T.G., Doerder,F.P., Jahn,C.L. and Herrick,G.				
JOURNAL	A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common 'D35E' motif				
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 91 (3), 942-946 (1994)				
AUTHORS	94134747				
TITLE	2 (bases 1 to 354)				
JOURNAL	Witherspoon,D.J., Doak,T.G., Williams,K., Seger,J. and Herrick,G.				
MEDLINE	Selection on the protein-coding genes of the TBE1 family of transposable elements in the ciliates Oxytricha fallax and O. trifallax				
AUTHORS	Unpublished				
TITLE	3 (bases 1 to 354)				
JOURNAL	Doak,T.G., Williams,K., Witherspoon,D.J. and Herrick,G.				
MEDLINE	Direct Submission				
AUTHORS	Submitted (11-FEB-1997) Oncological Science, University of Utah, School of Med. RmSC334, USA, UT 84132, USA				
TITLE	Location/Qualifiers				
JOURNAL	1. .354				
MEDLINE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
AUTHORS					
TITLE					
JOURNAL					

```
/organism="Oxytricha fallax"  
/strain="9D1"  
/db_xref="taxon:5944"  
/transposon="TBEI"  
/note=this is a bulk sequence that was generated from a PCR product that represents many transposon templates  
<!.>354  
/codon_start=1  
/transl_table=6  
/product="57kD zinc finger/protein chimera"  
/protein_id="AAB49844.1"  
/db_xref="pid:g1881676"  
/db_xref="GI:1881676"  
/translation="HYRDLKHLKAHKXEXEXXXLKLXLKRKAREXXXXXX  
QAXEVVXXNRNKLASEXTMKIKYKKIPVLQAIDLDTSLSQLILEDSDFKVKI  
XDHQAYKYFPFNL"
```

CDS		BASE COUNT	106 a	42 c	41 g	54 t	111 others	ORIGIN
-----	--	------------	-------	------	------	------	------------	--------

Query Match 3.0%: Score 34; DB 21; Length 354;
Best Local Similarity 26.5%; Pred. No. 1.13e-05;
Matches 36; Conservative 49; Mismatches 51; Indels 0; Gaps 0;

Db	41	ARAAAGADWTAGARRTTGARRAKHAWGMHRWARGRYTRAAAGYTDCCTCAAHAAAAGAARG 100
		: : : : : : : : : : :
Cp	765	agatatattgtagttccagaagcaagtaaatcagcatcacataactagagttaaag 706
		: : : : : : : : : : :
Db	101	CDAGRARYHHWDWARGWGYGCGYYKAAVAAGCWACMGARWASWTRYGTAAARGA 160
		: : ::::: : : : : : : : : : : : : : : : :
Cp	705	ataagaacatctccaggaaattaataatcagataatttacttatgcagtcgaagt 646
		: : ::::: : : : : : : : : : : : : : : : :
Db	161	WDKRCGCACAACCWWYT 176
		::: : : :::
Cp	645	tggcaatctcgcaatt 630
		:

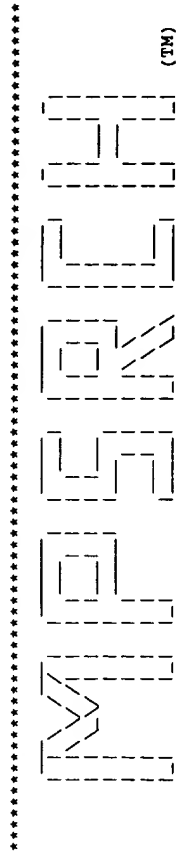
RESULT	11							
LOCUS	I28278	215 bp	DNA	PAT				30-OCT-1996
DEFINITION	Sequence 5 from patent US 5569830.							
ACCESSION	I28278							
NID	G1819054							
VERSION	I28278.1	GI:1819054						
KEYWORDS	Unknown.							
SOURCE	Unknown.							
ORGANISM	Unclassified. 1 (bases 1 to 215) Bennett,A., Labavitch,J.M., Powell,A. and Stotz,H. Plant inhibitors of fungal polygalacturonases and their use to control fungal disease Patent: US 5569830-A 5 29-OCT-1996; Location/Qualifiers 1..215 /organism="unknown"							
BASE COUNT	15 a	8 c	25 g	26 t	141 others			
ORIGIN								

Query Match 2.9%: Score 32; DB 25; Length 215;
Best Local Similarity 17.2%; Pred. No. 2.01e-04;
Matches 28; Conservative 63; Mismatches 70; Indels 2; Gaps 2;

Db	20	DKAKKDGNTHSSWTDCCNTGWCDFTT-YRVNNDSGHNKYSSANYNGNVGAARKT 78
		: : : ::: : : ::: : : ::: : : ::: : : :::
Qy	520	ttagatgttgattgacagattcagagattgagggaccgggttccaagaagaac 579
		: : : :::~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
Db	79	HXYHTHVNSGADSKTYVDYSNASGTSSSNGTGDNRSGADSYGSSKTAMTSRNPTCKTN 138
		: : : :::~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
Qy	580	tagaaagtgttgagcaatcatctgcggcagttggtggaactaatg-acgaaattgcgag 638
		: : : :::~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
Db	139	NAVDSRMWGDASVGSKNTKKHAKNSADKGKGVSKNNNGDRNNRY 181
		: : : :::~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
Qy	639	attgccaaacttcagctgctaagataattatctgatataat 681
		: : : :::~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:

```
RESULT 12
LOCUS OFU89262 370 bp DNA INV 14-MAR-1997
DEFINITION Oxytricha fallax 57kd zinc finger/protein kinase gene, partial cds.
ACCESSION U89262
NID 91881681
VERSION U89262.1 GI:1881681
KEYWORDS
SOURCE
ORGANISM Oxytricha fallax.
Eukaryotae; mitochondrial eukaryotes; Alveolata; Ciliophora;
hypotrichs; Stichotrichida; Oxytricha.
REFERENCE 1 (bases 1 to 370)
AUTHORS Doak,T.G., Doerder,F.P., Jahn,C.L. and Herrick,G.
TITLE A proposed superfamily of transposase genes: transposon-like
elements in ciliated protozoa and a common 'D35E' motif
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (3), 942-946 (1994)
MEDLINE 9413477
REFERENCE 2 (bases 1 to 370)
AUTHORS Witherspoon,D.J., Doak,T.G., Williams,K., Seger,J. and Herrick,G.
TITLE Selection on the protein-coding genes of the TBEl family of
transposable elements in the ciliates Oxytricha fallax and O.
trifallax
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 370)
AUTHORS Doak,T.G., Williams,K., Witherspoon,D.J. and Herrick,G.
TITLE Direct Submission
JOURNAL Submitted (11-FEB-1997) Oncological Science, University of Utah,
School of Med. Rm5C334, USA, UT 84132, USA
FEATURES
source
Location/Qualifiers
1..370
/organism="Oxytricha fallax"
/strain="9D1"
/db_xref="taxon:5944"
/transposon="TBEl"
/note="this is a bulk sequence that was generated from a
PCR product that represents many transposon templates"
<1..>370
/codon_start=3
/transl_table=6
/product="57kd zinc finger/protein kinase"
/protein_id="AAB49647.1"
/db_xref="PID:g1881682"
/db_xref="GI:1881682"
/translation="HTRDLXKLLKAKKXXEXXXXXXKLNKRNKAREXXXXXXXXX
OXAEYVXXRXNLXSEXTKIMIKIQYKIPVLAQIDLTSLQSLILEDSFDKVI
XDQYAKVYPENLXSDK"
BASE COUNT 116 a 43 c 48 g 69 t 94 others
ORIGIN
Query Match 2.9%; Score 33; DB 21; Length 370;
Best Local Similarity 33.1%; Pred. No. 4.81e-05;
Matches 45; Conservative 39; Mismatches 52; Indels 0; Gaps 0;
Db 43 ARAAGANWTAGARRTTGARRAKHAWCAMRHTSARYTGAGYDCTMAACAAGAAAG 102
Cp 765 agatatttctgttcccaagagcaagtaaatcagcatcagcatcagcttaag 706
Db 103 CRAGRGARYMTYDWARGBYCWGWCYTKAAYAAGCAGWAGHAGTAYGTDAAG 162
Cp 705 ataaagaacatctccaggaattaataatcagataatattctattatcagctgaag 646
Db 163 WDKACGCAACCCWMT 178
Cp 645 tggcaatctcgcaatt 630
RESULT 13
LOCUS AR024229 965 bp DNA PAT 04-DEC-1998
DEFINITION Sequence 22 from patent US 5795961.
ACCESSION AR024229
NID 93977523
VERSION AR024229.1 GI:3977523
```

```
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 965)
AUTHORS Wallace,T.Paul, Harris,W.J., Carr,F.J., Old,L.J., Welt,S. and
Kitamura,K.
TITLE Recombinant human anti-Lewis b antibodies
JOURNAL Patent: US 5795961-A 22 18-AUG-1998;
FEATURES Location/Qualifiers
source
1..965
/organism="unknown"
BASE COUNT 192 a 170 c 226 g 205 t 172 others
ORIGIN
Query Match 2.9%; Score 33; DB 25; Length 965;
Best Local Similarity 22.8%; Pred. No. 4.81e-05;
Matches 21; Conservative 45; Mismatches 24; Indels 2; Gaps 2;
Db 812 RGRGWGDYGGGYNYNGKRGRTVMTADTSNSRSSVTA-ADTAVYVYCVGRSVDSDGGDYW 870
Qy 823 gggcaagctgggtcattggaattaaaggaagattatgattgttcagaaaacgagta 882
Db 871 GGTTVTVSSHTVKDMTSSSSASVGDRTVTCRS 902
Qy 883 ggattacgataac-gcacagagaactgcac 913
RESULT 14
LOCUS HUAC004787 216021 bp DNA PRI 24-JUL-1998
DEFINITION Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10, complete
sequence.
ACCESSION AC004787
NID g3337381
VERSION AC004787.1 GI:3337381
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 216021)
AUTHORS Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
TITLE Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 216021)
AUTHORS Adams,M.D. and Loftus,B.J.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, Email:
b.loftus@tigr.org
REFERENCE 3 (bases 1 to 216021)
AUTHORS Adams,M.D. and Loftus,B.J.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
COMMENT On Jul 24, 1998 this sequence version replaced gi:3241936.
Address all correspondence to: Mark Adams The Institute for Genomic
Research 9712 Medical Center Dr., Rockville, MD 20850, USA e-mail
address: humgen@tigr.org. The orientation of the sequence is from
SP6 to T7 end. Genes were identified by a combination of five
methods including: XGRAIL (available by anonymous ftp from
arthur.epm.ornl.gov), Genefinder (Phil Green, University of
Washington), Genscan (Chris Burge,
http://genomic.stanford.edu/~chris/GENSCANW.html), searches of the
complete sequence against a peptide database, and the Human gene
Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
Genes without peptide homology having spliced EST hits are termed
'Unknown gene product'. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/CRNAscan-SE/).
FEATURES
source
Location/Qualifiers
1..216021
/organism="Homo sapiens"
```

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:00:35 1999; MasPar time 13.02 Seconds
Tabular output not generated. 807.364 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (1-372) from PCTUS9913024.pep (1 of 12)
Perfect Score: 2542
Sequence: 1 MEKFAEFGQGVQVPPFLSE.....LPTQASSKTKGVFGQEPREV 372

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 49.734; Variance 92.468; scale 0.538

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1893	74.5	295	1	TAIL FIBER PROTEIN GP3	0.00e+00
2	300	11.8	65	1	HYPOTHETICAL 7.3 KD PR	3.10e-35
3	124	4.9	196	1	4-HYDROXY-2-OXOSULFARA	7.12e-04
4	113	4.4	323	1	HEM2-HELPY	2.56e-02
5	108	4.2	672	1	NADH-UBIQUINONE OXIDOR	1.21e-01
6	106	4.2	673	1	HYPOTHETICAL 74.4 KD P	2.23e-01
7	106	4.2	2492	1	POLN-EEVVT	2.23e-01
8	103	4.1	154	1	OM22-NEUCR	5.46e-01
9	103	4.1	390	1	DHE4-SULSH	5.46e-01
10	103	4.1	723	1	REFX2-HUMAN	5.46e-01
11	105	4.1	727	1	NUAM-BOVIN	3.01e-01
12	105	4.1	727	1	NUAM-HUMAN	3.01e-01
13	105	4.1	1500	1	SSP5-STGRN	3.01e-01
14	104	4.1	2492	1	POLN-EEVVT	4.06e-01
15	99	3.9	421	1	DHE3-SULSO	1.75e-00
16	99	3.9	475	1	NCAP-SYNY	1.75e-00
17	98	3.9	699	1	YA76-SCHPO	2.32e+00
18	100	3.9	861	1	MUTS-HAEIN	1.31e+00
19	99	3.9	945	1	AMPE-MOUSE	1.75e+00
20	100	3.9	1300	1	POL2-MOUSE	1.31e+00
21	100	3.9	1693	1	POLN-HEVNV	1.31e+00
22	99	3.9	2492	1	POLN-EEVVT	1.75e+00
23	97	3.8	380	1	CAPM-STAAU	3.08e+00

24	97	3.8	542	1	OPPA_SALTY	PERIPLASMIC OLIGOPEPTI	3.08e-00
25	97	3.8	578	1	SHC_MOUSE	SHC TRANSFORMING PROTE	3.08e-00
26	97	3.8	583	1	SHC_HUMAN	SHC TRANSFORMING PROTE	3.08e-00
27	96	3.8	692	1	REFX2_MOUSE	DNA BINDING PROTEIN RF	4.08e-00
28	97	3.8	780	1	ACON_BOVIN	ACONITATE HYDRATASE, M	3.08e-00
29	96	3.8	833	1	GYRA_BACS5	DNA GYRASE SUBUNIT A (4.08e-00
30	96	3.8	969	1	MSUI_YEAST	MITOCHONDRIAL BIOGENES	4.08e-00
31	93	3.7	156	1	YEHS_ECOLI	HYPOTHETICAL 18.0 KD P	9.31e-00
32	95	3.7	171	1	YAL4_PHVU	HYPOTHETICAL AL4 PROTE	5.39e-00
33	95	3.7	196	1	YCH6_YEAST	HYPOTHETICAL 22.3 KD P	5.39e-00
34	94	3.7	303	1	ILID_ECOLI	HYPOTHETICAL ABC TRANS	7.09e-00
35	95	3.7	334	1	SYW_HAEIN	TRYPHOPHANYL-TRNA SYNT	5.39e-00
36	95	3.7	393	1	FLAA_PSEAE	FLAGELLIN.	5.39e-00
37	95	3.7	432	1	TIG_ECOLI	TRIGGER FACTOR (TF).	5.39e-00
38	95	3.7	499	1	NU4C_MARPO	NADH-PLASTOQUINONE OXI	5.39e-00
39	94	3.7	577	1	URE1_MYCTU	UREASE ALPHA SUBUNIT (7.09e-00
40	94	3.7	663	1	GLI3_CHICK	ZINC FINGER PROTEIN GL	7.09e-00
41	94	3.7	776	1	VP42_ROTTS	OUTER CAPSID PROTEIN V	7.09e-00
42	94	3.7	960	1	KKIT_CHICK	MAST/STEM CELL GROWTH	7.09e-00
43	94	3.7	997	1	YPX2_CAEEL	HYPOTHETICAL 110.3 KD	7.09e-00
44	94	3.7	1354	1	VGR2_HUMAN	VASCULAR ENDOTHELIAL G	7.09e-00
45	95	3.7	1958	1	UBRI_SCHPO	PROBABLE N-END-RECOGN	5.39e-00

ALIGNMENTS

RESULT 1
ID VG35_BPT4 STANDARD; PRT; 295 AA.
AC P03742;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE TAIL FIBER PROTEIN GP35.
GN 35.
OS BACTERIOPHAGE T4.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE;
OC T4-LIKE PHAGES.
RN [1]
RP SEQUENCE FROM N.A.
RA SELIVANOF N., REVEL H., MESYANZHINOV V.V., GOLDBER E.;
RL UNPUBLISHED OBSERVATIONS (XXX-1991).
RN [2]
RP SEQUENCE OF 238-295 FROM N.A.
RX MEDLINE: 82170495.
RA OLIVER D.B., CROWTHER R.A.;
RT "DNA sequence of the tail fibre genes 36 and 37 of bacteriophage T4.";
RL J. MOL. BIOL. 153:545-568(1991).
CC [1]-FUNCTION: STRUCTURAL COMPONENT OF THE DISTAL-HALF TAIL FIBER.
CC THE TAIL FIBER OF T4 IS ABOUT 1600 ANGSTROMS LONG WITH A KINK IN
CC THE MIDDLE THAT DIVIDES THE FIBER INTO PROXIMAL AND DISTAL HALVES.
CC THE THIN TIP OF THE DISTAL HALF-FIBER INTERACTS WITH THE BACTERIAL
CC LIPOPOLYSACCHARIDE RECEPTOR AND SPECIFIES THE HOST RANGE OF THE
CC PHAGE.
CC -1- SUBUNIT: THE DISTAL HALF-FIBER CONTAINS TWO MOLECULES EACH OF GP36
CC AND GP37 AND ONE MOLECULE OF GP35.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: V00863; G836604; -
CC EMBL: J02509; G215885; -
CC DR PIR: A04367; TLBP54.
CC DR PIR: JT0577; JT0577.
KW FIBER PROTEIN.
SQ SEQUENCE 295 AA; 32282 MW; 2F9ACC4A CRC32;

Query Match 74.5%; Score 1893; DB 1; Length 295;
Best Local Similarity 99.3%; Pred. No. 0.00e+00;

```
Matches 277; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 17 FMNSLTNNRIVAITLSKVNFPPEVSWLRTAGTSAPSDSILSRFDVSYAAPTSSKRA 76
Qy 94 FMNSLTNNRIVAITLSKVNFPPEVSWLRTAGTSAPSDSILSRFDVSYAAPTSSKRA 153
Db 77 IALSHVLSNRKSTDDYQOTILDVVFDSLEDVYGATGFPRTYSEVQFMSAVGGTNNETAR 136
Qy 154 IALSHVLSNRKSTDDYQOTILDVVFDSLEDVYGATGFPRTYSEVQFMSAVGGTNDIAR 213
Db 137 LPTSAISKSLDYNLIPGDVLYLKAQLYADADLLALGTNNISIRFYNASNGYISSTQAEF 196
Qy 214 LPTSAISKSLDYNLIPGDVLYLKAQLYADADLLALGTNNISIRFYNASNGYISSTQAEF 273
Db 197 TGOAGSWELKEDYVVPENAVGFTIYQRTAAGOGGMRNLSFSEVSRNGGISKPAEFGV 256
Qy 274 TGOAGSWELKEDYVVPENAVGFTIYQRTAAGOGGMRNLSFSEVSRNGGISKPAEFGV 333
Db 257 NGRVNVICESASPPDIWLPQTQASSKTGKVGQEFREV 295
Qy 334 NGRVNVICESASPPDIWLPQTQASSKTGKVGQEFREV 372
RESULT 2
ID Y15A_BP74 STANDARD; PRT; 65 AA.
AC P39509;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 7.3 KD PROTEIN IN GP34-GP35 INTERGENIC REGION.
GN Y15A OR 34.1.
OS BACTERIOPHAGE T4.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE;
OC T4-LIKE PHAGES.
RN [1]
RA KUTTER E.M.;
RL SUBMITTED (NOV-1994) TO THE SWISS-PROT DATA BANK.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 65 AA; 7334 MW; C2D7CE2D CRC32;
Query Match 11.8%; Score 300; DB 1; Length 65;
Best Local Similarity 82.1%; Pred. No. 3.10e-35;
Matches 46; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
Db 1 MEKFMAEIWTICPNAILSESRYKISTAGSCPLSTAGSPYVKFQDNVPGSQTF 56
Qy 1 MEKFMAEIWTICPNAILSESRYKISTAGSCPLSTAGSPYVKFQDNVPGSQTF 56
RESULT 3
ID ALKH_BACSU STANDARD; PRT; 196 AA.
AC P50846;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE 4-HYDROXY-2-OXOGLUTARATE ALDOLASE (EC 4.1.3.16) (2-KETO-4-
DE HYDROXYGLUTARATE ALDOLASE) (KHG-ALDOLASE) / 2-DEHYDRO-3-
DE DEOXYPHOSPHOGLUCONATE ALDOLASE (EC 4.1.2.14) (PHOSPHO-2-DEHYDRO-3-
DE DEOXYGLUCONATE ALDOLASE) (PHOSPHO-2-KETO-3-DEOXYGLUCONATE ALDOLASE)
DE (2-KETO-3-DEOXY-6-PHOSPHOGLUCONATE ALDOLASE) (KDPG-ALDOLASE).
GN KDA.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RA SEQUENCE FROM N.A.
RC STRAIN-168 / MARBURG;
RX MEDLINE; 96349105.
RA SOROKIN A.V., AZEVEDO V., ZUMSTEIN E., GALLERON N., EHRLICH S.D.,
RA SERROR P.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
RT the serA and kgd loci cloned in a yeast artificial chromosome.";
```

```
RL MICROBIOLOGY 142:2005-2016(1996).
CC -1- CATALYTIC ACTIVITY: 4-HYDROXY-2-OXOGLUTARATE - PYRUVATE +
CC GLYOXYLATE.
CC -1- CATALYTIC ACTIVITY: 2-DEHYDRO-3-DEOXY-D-GLUCONATE 6-PHOSPHATE -
CC PYRUVATE + D-GLYCERALDEHYDE 3-PHOSPHATE.
CC -1- PATHWAY: KEY ENZYME IN THE ENTNER-DOUDOROFF PATHWAY.
CC -1- PATHWAY: PARTICIPATES IN THE REGULATION OF THE INTRACELLULAR LEVEL
CC OF GLYOXYLATE.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE KHG/KDPG ALDOLASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL; L47838; G1146190; -
CC EMBL; Z99115; E1183656; -
CC DR SUBTILIS; BG11396; KDA.
CC DR PROSITE; PS00159; ALDOLASE_KDPG_KHG_1; FALSE_NEG.
CC DR PROSITE; PS00160; ALDOLASE_KDPG_KHG_2; 1.
CC DR PFAM; PF01081; Aldolase; 1.
CC KW LYASE; SCHIFF BASE; MULTIFUNCTIONAL ENZYME.
CC FT ACT_SITE 43 43 BY SIMILARITY.
CC FT ACT_SITE 47 47 BY SIMILARITY.
CC FT BINDING 130 130 SCHIFF-BASE WITH KHG OR PYRUVATE (BY
CC SIMILARITY).
CC SQ SEQUENCE 196 AA; 20865 MW; 14FCBED5 CRC32;
Query Match 4.9%; Score 124; DB 1; Length 196;
Best Local Similarity 27.8%; Pred. No. 7.12e-04;
Matches 20; Conservative 27; Mismatches 21; Indels 4; Gaps 4;
Db 114 TPSEIMEALTFGTTLKLFPSGVGIFPMKNLAGPFPQVTFIPGGIH-PSEVPDWLR-A 171
Qy 68 STGALVDSKSIATFSTNDTTSAAF-VSFMSNLNN-RIVALLTSGKVNFPPEVSWLRTA 125
Db 172 GAGAVGVGSQLG 183
Qy 126 GTSAPFSDSILS 137
RESULT 4
ID HEM2_HELPY STANDARD; PRT; 323 AA.
AC P56074;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DELTA-AMINOLEVULINIC ACID DEHYDRATASE (EC 4.2.1.24) (PORPHOBILINOGEN
DE SYNTHASE) (ALAD) (ALADH).
GN HEMB OR HP0163.
OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; HELICOBACTER GROUP;
OC HELICOBACTER.
RN [1]
RA SEQUENCE FROM N.A.
RC STRAIN-26695 / ATCC 700392;
RX MEDLINE; 97394467.
RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATHEY L., WALLIN E.,
RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA VENTER J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
```

RL NATURE 388:539-547(1997).
CC -1- CATALYTIC ACTIVITY: 2 5-AMINOLEVULINATE - PORPHOBILINOGEN +
CC 2 H(2)O.
CC -1- COFACTOR: ZINC.
CC -1- PATHWAY: SECOND STEP IN PORPHYRIN BIOSYNTHESIS.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: AE000537; G2313250; -
CC TIGR: HP0163; -
CC PROSITE: PS00169; D_ALA_DEHYDRATASE; 1.
CC PFAM: PF00490; ALAD; 1.
CC PORPHYRIN BIOSYNTHESIS; LYASE; ZINC.
CC DOMAIN 115 132 ZINC-BINDING (BY SIMILARITY).
CC ACT_SITE 246 246 BY SIMILARITY.
CC SEQUENCE 323 AA; 36202 MW; 9B60D720 CRC32;
CC
CC Query Match 4.4%; Score 113; DB 1; Length 323;
CC Best Local Similarity 25.7%; Pred. No. 2,56e-02;
CC Matches 27; Conservative 33; Mismatches 39; Indels 6; Gaps 6;
CC
CC Db 135 SVSNKDT-LEILNLOGLLAESGVDLIAPSNM-MDCNVLS-LRKTLDNAGYHTPTMS-Y 190
CC QY 81 STSNQTSAAFTSSKRAIALEHVKLSNRKSTD-DYQTLIDVVDLSLE 139
CC
CC Db 191 STKFASYGCPFRDVSANSPSFGDRKSYQMDYANQKALLESLED 235
CC QY 140 DVSYAFTSSKRAIALEHVKLSNRKSTD-DYQTLIDVVDLSLE 183
CC
CC RESULT 5
CC ID NQ03_PARDE STANDARD; PRT; 672 AA.
CC AC P29915;
CC DT 01-APR-1993 (REL. 25, CREATED)
CC DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
CC DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
CC DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 3 (EC 1.6.5.3) (NADH
CC DE DEHYDROGENASE 1, CHAIN 3) (NDH-1, CHAIN 3).
CC GN NQ03.
CC OS PARACOCUS DENITRIFICANS.
CC OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHODOBACTER GROUP;
CC OC PARACOCUS.
CC RN [1]
CC RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.
CC RC STRAIN-ATCC 13548;
CC RX MEDLINE; 92296779.
CC RA XU X., MATSUNO-YAGI A., YAGI T.;
CC RT "Structural features of the 66-kDa subunit of the energy-transducing
CC RT NADH-ubiquinone oxidoreductase (NDH-1) of Paracoccus denitrificans.";
CC RL ARCH. BIOCHEM. BIOPHYS. 296:40-48(1992).
CC RN [2]
CC RP SEQUENCE OF 657-672 FROM N.A.
CC RX MEDLINE; 93136200.
CC RA XU X., MATSUNO-YAGI A., YAGI T.;
CC RT "DNA sequencing of the seven remaining structural genes of the gene
CC RT cluster encoding the energy-transducing NADH-quinone oxidoreductase
CC RT of Paracoccus denitrificans.";
CC RL BIOCHEMISTRY 32:968-981(1993).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE - NAD(+) + UBIQUINOL.
CC -1- COFACTOR: MAY BIND TWO 4FE-4S CLUSTER AND ONE 2FE-2S CLUSTER.
CC -1- SUBUNIT: COMPOSED OF 14 DIFFERENT SUBUNITS.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 75 KD SUBUNIT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: M84572; G150601; -
CC PIR: S23948; S23948.
CC DR PROSITE: PS00641; COMPLEX1_75K_1; 1.
CC DR PROSITE: PS00642; COMPLEX1_75K_2; 1.
CC DR PROSITE: PS00643; COMPLEX1_75K_3; 1.
CC DR PFAM: PF00111; fer2; 1.
CC DR PFAM: PF00384; molybdopterln; 1.
CC KW OXIDOREDUCTASE; NAD; UBIQUINONE; IRON-SULFUR; 4FE-4S.
CC FT INIT_MET 0
CC FT METAL 25 25 IRON-SULFUR (2FE-2S) (POTENTIAL).
CC FT METAL 36 36 IRON-SULFUR (2FE-2S) (POTENTIAL).
CC FT METAL 47 47 IRON-SULFUR (2FE-2S) (POTENTIAL).
CC FT METAL 50 50 IRON-SULFUR (2FE-2S) (POTENTIAL).
CC FT METAL 109 109 IRON-SULFUR (2FE-2S) (POTENTIAL).
CC FT METAL 112 112 IRON-SULFUR (2FE-2S) (POTENTIAL).
CC FT METAL 118 118 IRON-SULFUR (2FE-2S) (POTENTIAL).
CC FT METAL 157 157 IRON-SULFUR (2FE-2S) (POTENTIAL).
CC FT METAL 160 160 IRON-SULFUR (4FE-4S) (POTENTIAL).
CC FT METAL 163 163 IRON-SULFUR (4FE-4S) (POTENTIAL).
CC FT METAL 207 207 IRON-SULFUR (4FE-4S) (POTENTIAL).
CC SQ SEQUENCE 672 AA; 73028 MW; 40DB761A CRC32;
CC
CC Query Match 4.2%; Score 108; DB 1; Length 672;
CC Best Local Similarity 26.5%; Pred. No. 1,21e-01;
CC Matches 22; Conservative 26; Mismatches 30; Indels 5; Gaps 5;
CC
CC Db 164 VRFTTEVAGITOMGTGRG-EDSEITSYLNQTLNMOGNIIDLCVPGLV-SKPVAFTA 221
CC QY 24 VRKISIAQCLSTAGPSYVVFQDNPVGSQTFSAGLHLRVFDP-STGALVDKSKYAFST 82
CC
CC Db 222 RPWELTKTESIDVMDALGSSIRI 244
CC QY 83 SN-DTTSAAVSEFMNSLTNN-RI 103
CC
CC RESULT 6
CC ID YA21_SYNY3 STANDARD; PRT; 673 AA.
CC AC P72929;
CC DT 15-DEC-1998 (REL. 37, CREATED)
CC DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
CC DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
CC DE HYPOTHETICAL 74.4 KD PROTEIN SLL1021.
CC GN SLL1021.
CC OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
CC OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE; 97061201.
CC RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
CC RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
CC RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
CC RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
CC RA TABATA S.;
CC RT "Sequence analysis of the genome of the unicellular cyanobacterium
CC RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
CC RT entire genome and assignment of potential protein-coding regions.";
CC RL DNA RES. 3:109-136(1996).
CC -1- SIMILARITY: TO B.SUBTILIS YUAG AND E.COLI YQIK.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: D90901; G1652020; -

[illegible]


```
DR PIR: A33552; A33552.
DR PROSITE; PS00641; COMPLEX1_75K_1; 1.
DR PROSITE; PS00642; COMPLEX1_75K_2; 1.
DR PROSITE; PS00643; COMPLEX1_75K_3; 1.
DR PFAM; PF00111; fer2; 1.
DR PFAM; PF00384; molybdopterin; 1.
KW OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION; TRANSIT PEPTIDE;
KW IRON-SULFUR; 4FE-4S.
FT TRANSIT 1 23 MITOCHONDRION.
FT CHAIN 24 727 NADH-UBIQUINONE REDUCTASE 75 KD SUBUNIT.
FT METAL 53 53 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 64 64 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 75 75 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 78 78 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 92 92 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 128 128 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 131 131 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 137 137 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 176 176 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 179 179 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 182 182 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 226 226 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ SEQUENCE 727 AA; 79442 MW; 9CAFI39C CRC32;

Query Match 4.1%; Score 105; DB 1; Length 727;
Best Local Similarity 28.6%; Pred. No. 3.01e-01;
Matches 26; Conservative 21; Mismatches 39; Indels 5; Gaps 5;

Db 183 IRFASEIAGVDDLTGTCGRN-DMOVGTIYIEKMFSELSGNIIIDICPVGALT-SKPYAFTA 240
QY 24 VRYKISAGSCPLSTAGPSVYKFDQNPVGSQTFSGALHLRVFDP-STGALVDSKSYAFST 82

Db 241 RPNWETRKTESIDVMDVAGSNIVVSTRT-GEV 270
QY 83 SN-DTTSAAAFVSPMNSLTNNRIVAILTSGKV 112

RESULT 12
ID NUAM_HUMAN STANDARD; PRT; 727 AA.
AC P28331;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DE 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE NADH-UBIQUINONE OXIDOREDUCTASE 75 KD SUBUNIT PRECURSOR (EC 1.6.5.3)
DE (EC 1.6.99.3) (COMPLEX I-75KD) (CI-75KD).
GN NDUFS1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX CHOW W., RAGAN I., ROBINSON B.H.;
RX MEDLINE; 92037608.
RA "Determination of the cDNA sequence for the human mitochondrial
RT 75-kDa Fe-S protein of NADH-coenzyme Q reductase.";
RL EUR. J. BIOCHEM. 201:547-550(1991).
CC -!- FUNCTION: THIS IS THE LARGEST SUBUNIT OF COMPLEX I AND IT IS A
CC COMPONENT OF THE IRON-SULFUR (IP) FRAGMENT OF THE ENZYME. IT
CC MAY FORM PART OF THE ACTIVE SITE CREVICE WHERE NADH IS OXIDIZED.
CC -!- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
CC -!- COFACTOR: MAY BIND ONE 2FE-2S CLUSTER & ONE 4FE-4S CLUSTER.
CC -!- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 30 DIFFERENT SUBUNITS.
CC -!- SUBCELLULAR LOCATION: MATRIX AND CYTOPLASMIC SIDE OF THE
CC MITOCHONDRIAL INNER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE COMPLEX I 75 KD SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; X61100; G38079; --
DR PIR: S16382; S16382.
DR PIR: S17854; S17854.
DR MIM; 157655; --
DR PROSITE; PS00641; COMPLEX1_75K_1; 1.
DR PROSITE; PS00642; COMPLEX1_75K_2; 1.
DR PROSITE; PS00643; COMPLEX1_75K_3; 1.
DR PFAM; PF00111; fer2; 1.
DR PFAM; PF00384; molybdopterin; 1.
KW OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION; TRANSIT PEPTIDE;
KW IRON-SULFUR; 4FE-4S.
FT TRANSIT 1 23 MITOCHONDRION (BY SIMILARITY).
FT CHAIN 24 727 NADH-UBIQUINONE REDUCTASE 75 KD SUBUNIT.
FT METAL 53 53 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 64 64 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 75 75 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 78 78 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 92 92 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 128 128 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 131 131 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 137 137 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 176 176 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 179 179 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 182 182 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 226 226 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ SEQUENCE 727 AA; 79573 MW; B9F33158 CRC32;

Query Match 4.1%; Score 105; DB 1; Length 727;
Best Local Similarity 28.6%; Pred. No. 3.01e-01;
Matches 26; Conservative 21; Mismatches 39; Indels 5; Gaps 5;

Db 183 IRFASEIAGVDDLTGTCGRN-DMOVGTIYIEKMFSELSGNIIIDICPVGALT-SKPYAFTA 240
QY 24 VRYKISAGSCPLSTAGPSVYKFDQNPVGSQTFSGALHLRVFDP-STGALVDSKSYAFST 82

Db 241 RPNWETRKTESIDVMDVAGSNIVVSTRT-GEV 270
QY 83 SN-DTTSAAAFVSPMNSLTNNRIVAILTSGKV 112

RESULT 13
ID SSP5_STRGN STANDARD; PRT; 1500 AA.
AC P16952; Q54184;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE AGGLUTININ RECEPTOR PRECURSOR.
DE SSP5 OR SSPB.
OS STREPTOCOCCUS GORDONII.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M5;
RX MEDLINE; 90236997.
RX DEMUTH D.R., GOLUB E.E., MALAMUD D.;
RT "Streptococcal-host interactions. Structural and functional analysis
RT of a Streptococcus sanguis receptor for a human salivary
RT glycoprotein.";
RL J. BIOL. CHEM. 265:7120-7126(1990).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RC STRAIN-M5;
RX MEDLINE; 96310377.
RX DEMUTH D.R., DUAN Y., BROOKS W., HOLMES A.R., MCNAB R.,
RA JENKINSON H.F.;
RT "Tandem genes encode cell-surface polypeptides SspA and SspB which
RT mediate adhesion of the oral bacterium Streptococcus gordonii to
RT human and bacterial receptors.";
RL MOL. MICROBIOL. 20:403-413(1996).
CC -!- FUNCTION: MAY BIND SALIVARY ACID RESIDUES OF SALIVARY AGGLUTININ
CC (SAG) IN A CALCIUM-DEPENDENT REACTION. THE INTERACTION OF SAG WITH
```



```

CC ITS RECEPTOR IN VARIOUS ORAL STREPTOCOCCI MODULATE BACTERIAL
CC COLONIZATION OF ORAL TISSUE AND IS ASSOCIATED WITH REDUCED LEVELS
CC OF DENTAL CARIES.
CC
CC -!- DOMAIN: THE PR2 REGION, BY SIMILARITY WITH THE PROLINE RICH
CC DOMAINS OF THE S PYOGENES M6 PROTEIN AND STAPHYLOCOCCAL PROTEIN A,
CC MAY TRAVERSE THE CELL WALL PEPTIDOGLYCAN AND IS FOLLOWED BY
CC HYDROPHOBIC AA WHICH MAY FUNCTION TO ANCHOR THE PROTEIN TO THE
CC MEMBRANE.
CC
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, CELL WALL.
CC
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC
CC -!- SIMILARITY: BELONGS TO THE SPAP/SSP5/SPAA FAMILY.
CC
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL: U40026; GI100973; -.
CC
CC PTR: A35186; A35186.
CC
CC DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
CC
CC KW SIGNAL; REPEAT: CALCIUM-BINDING; TRANSMEMBRANE.
CC
CC FT SIGNAL 1 38
CC FT CHAIN 39 1500
CC FT DOMAIN 39 1474
CC FT TRANSMEM 1475 1495
CC FT DOMAIN 1496 1500
CC FT SIMILAR 164 470
CC FT DOMAIN 164 470
CC FT REPEAT 242 323
CC FT REPEAT 242 323
CC FT REPEAT 324 405
CC FT REPEAT 406 470
CC FT DOMAIN 771 887
CC FT DOMAIN 1414 1436
CC FT FT DOMAIN 1467 1472
CC
CC CA_BIND 220 235
CC CA_BIND 301 316
CC FT CA_BIND 931 950
CC FT CA_BIND 1300 1315
CC
CC SQ SEQUENCE 1500 AA; 164552 MW; 9361C26C CRC32;
CC
CC
CC Query Match 4.18; Score 105; DB 1; Length 1500;
CC Best Local Similarity 25.08; Pred. No. 3.01e-01;
CC Matches 25; Conservative 32; Mismatches 37; Indels 6; Gaps 6;
CC
CC
CC Db 352 NEATKORNETAKATYDAAVKKYEADLAAVKQANATNADYQAKLAA-YOTELARVQKANA 410
CC QY 132 SDSILSRFDVSYAAFYTSRRK-AIALEHVKLSRKSTDDYQTTILDVVVDS-LEDVG-ATG 188
CC
CC
CC Db 411 DAKATYKAVEDNKKAKNAATKAENEETKQRNAVAK-TDYE 449
CC QY 189 FPRGTYE-SVEQFMASVGGTNDTEARLPTSAATKLSLDYN 227
CC
CC
CC RESULT 14
CC ID POLN_EEVV3 STANDARD; PRT: 2492 AA.
CC AC P36327;
CC DT 01-JUN-1994 (REL. 29, CREATED)
CC DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
CC DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
CC DE NONSTRUCTURAL POLYPROTEIN [CONTAINS: NONSTRUCTURAL PROTEINS NSP1 TO
CC NSP4].
CC OS VENEZUELAN EQUINE ENCEPHALITIS VIRUS (STRAIN 3880).
CC OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; TOGAVIRIDAE;
CC ALPHAVIRUS.
CC
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE; 93079859.

```



```
FT MOD_RES 1 1 ACETYLATION.
FT MOD_RES 254 254 METHYLATION (MONO-).
FT MOD_RES 260 260 METHYLATION (MONO-).
FT MOD_RES 372 372 METHYLATION (MONO-).
FT MOD_RES 391 391 METHYLATION (MONO-).
FT MOD_RES 392 392 METHYLATION (MONO-).
FT MOD_RES 393 393 METHYLATION (MONO-).
FT ACT_SITE 108 108 BY SIMILARITY.
SQ SEQUENCE 421 AA; 46157 MW; 2BE7B538 CRC32;

Query Match 3.9%; Score 99; DB 1; Length 421;
Best Local Similarity 21.7%; Pred. No. 1.75e+00;
Matches 15; Conservative 24; Mismatches 27; Indels 3; Gaps 3;

Db 122 EEOISRKYIOAIYKYLGSGLDIPAPDVNTDSOTMAWFL-DEYIKITGK-VDFAVETGKP 179
QY 245 DLLALGTTNISIRF-YNASNGYISSTQAEFTGQAGSWELAKEDYVVPENAVGFTIYAQT 303

Db 180 VELGGIGVR 188
QY 304 AQAGGGGMR 312
```

Search completed: Wed Sep 1 16:00:59 1999
Job time : 24 secs.

W E S L E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:02:42 1999; MasPar time 18.01 Seconds
Tabular output not generated. 827.821 Million cell updates/sec

Title: >PCT-US99-13024-2

Description: (1-372) from PCTUS9913024.pap (1 of 12)

Perfect Score: 2542

Sequence: 1 MEKFMAEFGGQVGTPTFLSE.....LPTQASSKTKGVFGQEFREV 372

Scoring table: PAM 150

Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: pir60

1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 48.649; Variance 104.828; scale 0.454

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1464	57.6	275	1	TLBP54	tail fiber protein gp	4.61e-250
2	128	5.0	510	2	S35028	protein p2 precursor	1.88e-03
3	124	4.9	196	2	H69647	deoxyphosphogluconate	6.17e-03
4	113	4.4	323	2	C64540	delta-aminolevulinic	1.45e-01
5	109	4.3	323	2	H71968	delta-aminolevulinic	4.36e-01
6	108	4.2	673	2	A45456	NADH dehydrogenase (u	5.72e-01
7	106	4.2	673	2	S74795	hypothetical protein	9.79e-01
8	105	4.2	2492	1	MNWVTD	nonstructural polypro	9.79e-01
9	103	4.1	154	1	A40669	mitochondrial recepto	2.16e+00
10	103	4.1	270	2	B71160	probable methyl-accep	2.16e+00
11	103	4.1	300	2	S37407	glutamate dehydrogena	2.16e+00
12	103	4.1	723	2	B55926	DNA binding protein R	2.16e+00
13	105	4.1	727	2	S17854	NADH dehydrogenase (u	1.28e+00
14	105	4.1	727	2	A33552	NADH dehydrogenase (u	1.28e+00
15	105	4.1	1473	2	A35186	salivary agglutinin r	1.28e+00
16	104	4.1	2492	1	C44213	nonstructural polypro	1.66e+00
17	102	4.0	119	2	B32578	T-cell receptor beta	2.81e+00
18	102	4.0	321	2	H71729	hypothetical protein	2.81e+00
19	101	4.0	548	2	S52621	adenosine phosphoribosyltr	3.64e+00
20	101	4.0	1004	2	D71490	probable exodeoxyribo	3.64e+00
21	98	3.9	311	2	S66469	catechol 1,2-dioxygen	7.82e+00
22	99	3.9	421	2	S20286	glutamate dehydrogena	6.07e+00
23	99	3.9	475	1	VHVNYS	nucleoprotein - Sonch	6.07e+00

24	98	3.9	572	2	S77249	hypothetical protein	7.82e+00
25	98	3.9	699	2	S62408	hypothetical protein	7.82e+00
26	100	3.9	861	2	G64087	DNA mismatch repair p	4.70e+00
27	99	3.9	945	2	S30398	aminopeptidase N homo	6.07e+00
28	99	3.9	1007	2	PN0156	glutamate receptor ch	6.07e+00
29	99	3.9	1008	2	S28858	glutamate receptor de	6.07e+00
30	100	3.9	1281	1	GNMSLL	retrovirus-related re	4.70e+00
31	99	3.9	2492	1	A44213	nonstructural polypro	6.07e+00
32	98	3.9	2567	2	A49551	filamin, Mueller cell	7.82e+00
33	99	3.9	2902	2	C71953	toxin-like outer memb	6.07e+00
34	96	3.8	163	2	D71169	hypothetical protein	1.29e+01
35	97	3.8	327	2	A60474	hypothetical protein	1.00e+01
36	97	3.8	459	2	A55484	p52(Shc) protein - mo	1.00e+01
37	97	3.8	473	2	S25776	transforming protein	1.00e+01
38	97	3.8	542	1	QREBOA	oligopeptide-binding	1.29e+01
39	96	3.8	586	2	PC6006	scaffolding protein C	1.29e+01
40	96	3.8	692	2	C55926	DNA binding protein R	1.29e+01
41	96	3.8	969	2	S54594	MSU1 protein - yeast	1.29e+01
42	96	3.8	1185	2	H64526	delta-1-pyrroline-5-c	1.29e+01
43	96	3.8	1185	2	B71980	proline/pyrroline-5-c	1.29e+01
44	96	3.8	1193	2	S68218	botulinum neurotoxin	1.29e+01
45	95	3.7	394	2	A37853	flagellin, 40K - Pseu	1.65e+01

ALIGNMENTS

RESULT 1

ENTRY TLBP54 #type complete
TITLE tail fiber protein gp35 - phase T4
ORGANISM #formal_name phase T4
DATE 06-Jul-1982 #sequence_revision 19-Apr-1996 #text_change 29-May-1998

ACCESSIONS JT0577: A04367
REFERENCE JT0576
#authors Mesyanzhinov, V.V.
#submission submitted to JIPID, April 1991
#accession JT0577
#status preliminary

#molecule_type protein
#residues 1-275 #label MES

REFERENCE A92877
#authors Oliver, D.B.; Crowther, R.A.
#journal J. Mol. Biol. (1981) 153:545-568
#title DNA sequence of the tail fibre genes 36 and 37 of bacteriophage T4.
#cross-references MUID:82170495.
#accession A04367

#molecule_type DNA
#residues 218-275 #label OLI
#note the gene gp35 product was identified on the basis of its map position

GENETICS

#gene gp35
#map_position 91.6-92.3

COMPLEX the distal half-fiber contains two molecules each of gp36 (PIR:TLBP64) and gp37 (PIR:TLBP74), and one molecule of gp35

CLASSIFICATION

#superfamily phase T4 tail fiber protein gp35
KEYWORDS structural protein; tail fiber

SUMMARY #length 275 #molecular-weight 30237 #checksum 7602

Query Match 57.6%; Score 1464; DB 1; Length 275;

Best Local Similarity 87.5%; Pred. No. 4.61e-250;

Matches 244; Conservative 11; Mismatches 4; Indels 20; Gaps 1;

Db 17 FNNSLTNNRIVAILTSGKNFPPVVSUWLRAGT-----TSSKRA 56

QY 94 FNNSLTNNRIVAILTSGKNFPPVVSUWLRAGT-----TSSKRA 153

Db 57 IALEHVKLSNRKSTDDYQIILWVDSLDVGATGPPRTYVESVEFMGAVGTNNIEAR 116

QY 154 IALEHVKLSNRKSTDDYQIILWVDSLDVGATGPPRTYVESVEFMGAVGTNNIEAR 213

Db 117 LPTSAATSKSDYNLIPGDVLYLKAQYADADLLALGTTNISIRFYNASNGYISSTQAEF 176
QY 214 LPTSAATSKSDYNLIPGDVLYLKAQYADADLLALGTTNISIRFYNASNGYISSTQAEF 273
Db 177 TGOAGSWELKEDYVLLPQNLPFTYHPHTPPQPGGMRNLSPFSEVRNGGISPAEFGV 236
QY 274 TGOAGSWELKEDYVVVFVNVAVGFTYIAQRTAQAGGMRNLSPFSEVRNGGISPAEFGV 333
Db 237 NGIRVNTICESASPPDLMVLPTQASSKTGKVGQEFREV 275
QY 334 NGIRVNTICESASPPDLMVLPTQASSKTGKVGQEFREV 372
RESULT 2
ENTRY #type complete
TITLE protein PS2 precursor - Corynebacterium glutamicum
ORGANISM #formal_name Corynebacterium glutamicum
DATE 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 17-Mar-1999
ACCESSIONS S35028
REFERENCE S35028
#authors Peyret, J.L.; Bavan, N.; Joliff, G.; Gulik-Krzywicki, T.; Mathieu, L.; Shechter, E.; Leblon, G.
#journal Mol. Microbiol. (1993) 9:97-109
#title Characterization of the cspB gene encoding PS2, an ordered surface-layer protein in Corynebacterium glutamicum.
#cross-references MUID:94018618
#accession S35028
#molecule_type DNA
##residues 1-510 #label PEY
##cross-references EMBL:X69103; NID:g40488; PID:g40489
##note the authors translated the codon AAC for residue 155 as Met
GENETICS cspB
KEYWORDS glycoprotein
FEATURE
1-30
31-510 #domain signal sequence #status predicted #label SIG\
38,58,80,187,332, #product protein PS2 #status predicted #label MAT\
353,469 #binding_site carbohydrate (Asn) (covalent) #status predicted
SUMMARY #length 510 #molecular-weight 55425 #checksum 3744
Query Match 5.0%; Score 128; DB 2; Length 510;
Best Local Similarity 23.3%; Pred. No. 1.88e-03;
Matches 52; Conservative 55; Mismatches 100; Indels 16; Gaps 13;
Db 134 ESNAYALKVDQEAAT-AFAEAYRNALRDAAI-SINPDGGIN-PDTSINLLIDAANAARND 190
QY 74 DSKSYAFSTNDTTSAAAFVSMNSLTNNRIVAILTSGKVNFPPEVVSMLRTAGTSAPPSD 133
Db 191 R--AEIE-DYAHLYTOT--DIALETPLQAYAFQDLKALQAEVDADFELGFEFGIDQ-EDG 244
QY 134 SILSRFDVSAAYFTSKRAIALEHVKLSNR-KSTDYQIILDVDFDSLEDVGATGPPRG 192
Db 245 NY--VORYHLPVAVEALKAEDVARVAATEPLRADSIKAKLSQSDVLVRLFLERATAQR 302
QY 193 TYESVEQF-MSAVGGTNDIEI-ARLPTSAAISKLS-DYNLIPGDVLYLKAQYADADLLAL 249
Db 303 DTLRVVEAIFSTARYVELYENNVNENKTLRQHSALIPN 345
QY 250 GTTNISIRFYNASNGYISSTQAEFTGOAGSWELKEDY-VVYPE 291
RESULT 3
ENTRY #type complete
TITLE deoxyphosphoglucuronate aldolase kdgA - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998
ACCESSIONS H69647
REFERENCE A69580

#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conner, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Erington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.; Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
#cross-references MUID:98044033
#accession H69647
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
##residues 1-196 #label KUN
##cross-references GB:299115; GB:AL009126; NID:g2634478; PID:el183656; PID:g2634629
##experimental_source strain 168
GENETICS kdgA
#gene
CLASSIFICATION #superfamily 2-dehydro-3-deoxyphosphoglucuronate aldolase
SUMMARY #length 196 #molecular-weight 20865 #checksum 1121
Query Match 4.9%; Score 124; DB 2; Length 196;
Best Local Similarity 27.8%; Pred. No. 6.17e-03;
Matches 20; Conservative 27; Mismatches 21; Indels 4; Gaps 4;
Db 114 TPSEIMEALTFGTTLLKLPSPGVGFGPEKMLAGPQVTFITFGGTH-PSEVPDMLR-A 171
QY 68 STGALVDSKISAEFSTNDTTSAAAF-VSEFNSMTNN-RIVAILTSGKVNFPPEVVSMLRTA 125
Db 172 GAGAVGVGSQGL 183
QY 126 GTSAPFSDSILS 137
RESULT 4
ENTRY #type complete
TITLE delta-aminolevulinic acid dehydratase - Helicobacter pylori (strain 26695)
ORGANISM #formal_name Helicobacter pylori
DATE 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 31-Oct-1997
ACCESSIONS C64540

```

REFERENCE
#authors
A64520
Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.;
McKenney, K.; Fitzegerald, L.M.; Lee, N.; Adams, M.D.;
Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
Fuji, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
C.M.; Venter, J.C.
#journal
Nature (1997) 388:539-547
#title
The complete genome sequence of the gastric pathogen
Helicobacter pylori.
#cross-references
MUID:97394467
#accession
C64540
#status
preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type
DNA
#residues
1-323 ##label TOM
#cross-references
GB:AE000537; GB:AE000511; NID:g2313247; PID:g2313250;
TIGR:HP0163
CLASSIFICATION
#superfamily porphobilinogen synthase
SUMMARY
#length 323 #molecular-weight 36202 #checksum 1005
Query Match
Best Local Similarity 4.4%; Score 113; DB 2; Length 323;
Matches 27; Conservative 33; Mismatches 39; Indels 6; Gaps 6;
Db 135 SVSNDKT-LEILNLOGLIAESVDILAPSNM-MDGNVLS-LRKTLDNAGYHTPTMS-Y 190
QY 81 STSNDTTSFAFVSFMSLNNRIVALTSGKNFPEVSWLRTA-GTSAFSDSILSRF 139
Db 191 STKFASYYGPPRDVANSAPSGDRKSYMDYANQKALELSELD 235
QY 140 DVSAAFYSSKRAIALEHVKLSNRKSTD-DYQTILDVVDLSLED 183

RESULT 5
ENTRY H71968 #type complete
TITLE delta-aminolevulinic acid dehydratase - Helicobacter pylori
ORGANISM (strain J99)
#variety #formal_name Helicobacter pylori
DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
05-Mar-1999
ACCESSIONS H71968
REFERENCE A71800
#authors Alm, R.A.; Ling, L.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.;
Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
G.F.; Trust, T.J.
#journal
Nature (1999) 397:176-180
#title
Genomic sequence comparison of two unrelated isolates of the
human gastric pathogen Helicobacter pylori.
#cross-references
MUID:99120557
#accession H71968
#status preliminary
#molecule_type DNA
#residues 1-323 ##label ARN
#cross-references
GB:AE001453; GB:AE001439; NID:g4154651; PID:g4154663
#experimental_source strain J99
GENETICS
#gene hmbB
CLASSIFICATION
#superfamily porphobilinogen synthase
SUMMARY
#length 323 #molecular-weight 36251 #checksum 2807
Query Match
Best Local Similarity 4.3%; Score 109; DB 2; Length 323;
Matches 23; Conservative 26; Mismatches 28; Indels 5; Gaps 5;

Db 157 VDILAPSNM-MDGNVLS-LRKALDKAGYHTPTMS-YSTKFASYYGPPRDVANSPPSFG 213
QY 104 VALTSGKNFPEVSWLRTA-GTSAFSDSILSRFDVSYAAFYSSKRAIALEHVKLS 162
Db 214 DRKSYMDYANQKALELSELD 235
QY 163 NRKSTD-DYQTILDVVDLSLED 183

RESULT 6
ENTRY A45456 #type complete
TITLE NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain N003 -
Paracoccus denitrificans
ORGANISM #formal_name Paracoccus denitrificans
DATE 24-Feb-1994 #sequence_revision 12-Apr-1996 #text_change
04-Sep-1998
ACCESSIONS S23948; A45456
REFERENCE S23946
#authors Xu, X.; Matsuno-Yagi, A.; Yagi, T.
#journal Arch. Biochem. Biophys. (1992) 296:40-48
#title Structural features of the 66-kDa subunit of the
energy-transducing NADH-ubiquinone oxidoreductase (NDH-1)
of Paracoccus denitrificans.
#cross-references
MUID:92296779
#accession S23948
#status preliminary
#molecule_type DNA
#residues 1-673 ##label XUA
#cross-references
GB:M84572; NID:g150599; PID:g150601
REFERENCE A45456
#authors Xu, X.; Matsuno-Yagi, A.; Yagi, T.
#journal Biochemistry (1993) 32:968-981
#title DNA sequencing of the seven remaining structural genes of the
gene cluster encoding the energy-transducing NADH-quinone
oxidoreductase of Paracoccus denitrificans.
#cross-references
MUID:93136200
#accession A45456
#status preliminary
#molecule_type DNA
#residues 658-673 ##label XU1
#note sequence extracted from NCBI backbone (NCBIN:123409,
NCBIP:123410)
CLASSIFICATION
#superfamily NADH dehydrogenase (ubiquinone) chain 11
KEYWORDS NAD; oxidoreductase
SUMMARY #length 673 #molecular-weight 73159 #checksum 4307
Query Match
Best Local Similarity 4.2%; Score 108; DB 2; Length 673;
Matches 22; Conservative 26; Mismatches 30; Indels 5; Gaps 5;
Db 165 VRFTEVAGITONGTGRC-EDSEITSYLNQTLSENMQGNIDLCVPGLV-SKPYAFTA 222
QY 24 VYKTSIAGSCPLSTAGSYVRFQDNPVGSQTFSGALHLRVFDP-STGALVDSKSYAFST 82
Db 223 RPWELTKTESIDVMDALGSSIRI 245
QY 83 SN-DTTSAAFFVSFMSLTNN-RI 103

RESULT 7
ENTRY S74795 #type complete
TITLE hypothetical protein sll1021 - Synechocystis sp. (strain PCC
6803)
ORGANISM #formal_name Synechocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
21-Aug-1998
ACCESSIONS S74795
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;

```

```

Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
DNA Res. (1996) 3:109-136
#journal
#title
Sequence analysis of the genome of the unicellular
cyanobacterium Synechocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references MUID:97061201
#accession S74795
##status preliminary
##molecule_type DNA
##residues 1-673 #label KAN
##cross-references EMBL:D90901; GB:AB001339; NID:g1651897; PID:d1017679;
PID:g1652020
##note
the nucleotide sequence was submitted to the EMBL Data
Library, June 1996
SUMMARY
#length 673 #molecular-weight 74423 #checksum 9140
Query Match 4.2%; Score 106; DB 2; Length 673;
Best Local Similarity 20.3%; Pred. No. 9.79e-01;
Matches 13; Conservative 26; Mismatches 21; Indels 4; Gaps 4;
Db 126 SLREISIDVVRAGNLAVRTQDYMRANRVYVCITPNRNEILTAAARLS-KKGQISEA 183
QY 310 GMRNLSEVSRNGGIS-KPAEFGVNGIRVN-YICESASPPDMVLPTQASSKTGVFGQ 367
Db 184 DIKD 187
QY 368 EFRE 371
RESULT 8 MNWTD #type complete
ENTRY nonstructural polyprotein - Venezuelan equine encephalitis
TITLE virus (strain TRD)
CONTAINS nonstructural protein NS1; nonstructural protein NS2;
nonstructural protein NS3; nonstructural protein NS4
ORGANISM #formal_name Venezuelan equine encephalitis virus
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
08-Apr-1994
ACCESSIONS A31467
REFERENCE A31467
#authors Kinney, R.M.; Johnson, B.J.B.; Welch, J.B.; Tsuchiya, K.R.;
Trent, D.W.
#journal Virology (1989) 170:19-30
#title The full-length nucleotide sequences of the virulent Trinidad
donkey strain of Venezuelan equine encephalitis virus and
its attenuated vaccine derivative, strain TC-83.
#cross-references MUID:89243175
#accession A31467
##molecule_type mRNA
##residues 1-2492 #label KIN
##note readthrough of the terminator UGA between codons CAA for
1879-Gln and CGG for 1880-Arg occurs
CLASSIFICATION #superfamily Semliki Forest virus nonstructural protein
KEYWORDS nonstructural protein; polyprotein
FEATURE
1-535 #product nonstructural protein NS1 #status predicted
#label NS1\
536-1329 #product nonstructural protein NS2 #status predicted
#label NS2\
1330-1879 #product nonstructural protein NS3 #status predicted
#label NS3\
1880-2492 #product nonstructural protein NS4 #status predicted
#label NS4\
SUMMARY #length 2492 #molecular-weight 277902 #checksum 8861
Query Match 4.2%; Score 106; DB 1; Length 2492;
Best Local Similarity 24.7%; Pred. No. 9.79e-01;
Matches 23; Conservative 26; Mismatches 39; Indels 7; Gaps 7;

```

```

Db 1125 LRNYDPRINDVPNRRLPHALVLHNEHPOSDFSSFSKL-KGRTVLVV-GEKLSVPGKM 1182
QY 62 LRVEDFSTGAL-VDSK-SYAFSTS-NDTTSAAVFSFNLSLNNRIYAILISGKVNFPPEV 118
Db 1183 VDWLSDRPEATFRARLDLGIPIGVKPYDIIFVNVNR 1217
QY 119 VSWLRAGTSAFPSDSILS-REDVS-YAAFYTSK 151
RESULT 9 A40669 #type complete
ENTRY mitochondrial receptor complex chain MOM22 - Neurospora
TITLE crassa
ALTERNATE_NAMES mitochondrial outer membrane preprotein receptor chain MOM22
ORGANISM #formal_name Neurospora crassa
DATE 03-May-1994 #sequence_revision 02-Jun-1994 #text_change
05-Sep-1997
ACCESSIONS A40669; S33472
REFERENCE A40669
#authors Kiebler, M.; Keil, P.; Schneider, H.; van der Klei, I.J.;
Pfanner, N.; Neupert, W.
#journal Cell (1993) 74:483-492
#title The mitochondrial receptor complex: a central role of MOM22
in mediating preprotein transfer from receptors to the
general insertion pore.
#cross-references MUID:93351229
#accession A40669
##molecule_type mRNA
##residues 1-154 #label KIE
##cross-references GB:X71021; NID:g311426; PID:g311427
COMMENT This protein resides in the mitochondrial outer membrane (MOM)
where it serves as part of the receptor complex for mitochondrial
preproteins.
GENETICS
#gene MOM22
CLASSIFICATION #superfamily mitochondrial receptor complex chain MOM22
KEYWORDS mitochondrion; transmembrane protein
FEATURE
85-105 #domain transmembrane #status predicted #label TMM
SUMMARY #length 154 #molecular-weight 16816 #checksum 9898
Query Match 4.1%; Score 103; DB 1; Length 154;
Best Local Similarity 28.2%; Pred. No. 2.16e+00;
Matches 37; Conservative 33; Mismatches 48; Indels 13; Gaps 12;
Db 10 EHFQQQVQVGEDEFTDTSISVDSYE--SQET-F-TDRLYALRDMVSTPTGRWFY 65
QY 198 EQPM-SAVGQTNDEIARLPTSAISKLSIDYNIPIGDVLYLKAOLYADADLLALGTTNISI 256
Db 66 HKYSTTTNFVKSTLS-FAGRA-AWAVSVSGLLIGVP-FATAFAE-DONYAAMEQEARMRE 121
QY 257 RFYNASNGVISSTQAEFTGAGSWELK-EDYVV-VPENAVGFTIYAQRTAQAGOGG-MRN 313
Db 122 LG-SDVLTAGG 131
QY 314 LSPSEVSRNGG 324
RESULT 10 B71160 #type complete
ENTRY probable methyl-accepting chemotaxis protein - Pyrococcus
TITLE horikoshii
ORGANISM #formal_name Pyrococcus horikoshii
DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
14-Aug-1998
ACCESSIONS B71160
REFERENCE A71000
#authors Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;

```

```
#journal      Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
#title        DNA Res. (1998) 5:55-76
#             Complete sequence and gene organization of the genome of a
#             hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
#             OT3.
#cross-references MUID:98344137
#accession     B71160
#status        preliminary: nucleic acid sequence not shown;
#              translation not shown
#molecule_type DNA
#residues      1-277 #label KAW
#cross-references GB:AP000002; NID:g3236129; PID:d1030510; PID:g3256884
#experimental_source strain OT3
#note          This accession replaces an interim accession for a
#              sequence replaced by GenBank

GENETICS
#gene          PH0479
SUMMARY
#length 277 #molecular-weight 30608 #checksum 4645

Query Match      4.1%; Score 103; DB 2; Length 277;
Best Local Similarity 28.6%; Pred. No. 2.16e+00;
Matches 16; Conservative 20; Mismatches 19; Indels 1; Gaps 1;

Db 77 ETSR-EMSNATILAEYSKKIDEIVYVISSIASQTNLLNLSIEARAGEGRGF 131
QY 210 EIALPTSAISKSDYNLPDGVLYLKAQLYADADLLAGTTNISIRFYNASNGY 265

RESULT 11
ENTRY   S37407 #type fragment
TITLE   glutamate dehydrogenase (NADP+) (EC 1.4.1.4) - Sulfolobus
        shibatae (fragment)
ORGANISM #formal_name Sulfolobus shibatae
DATE     09-Dec-1993 #sequence_revision 10-Nov-1995 #text_change
        25-Mar-1998
ACCESSIONS S37407
REFERENCE   S37407
#authors   Benachenhou-Lahfa, N.
#submission submitted to the EMBL Data Library, July 1993
#accession S37407
#molecule_type DNA
#residues  1-390 #label BEN
#cross-references EMBL:X73990
CLASSIFICATION #superfamily leucine dehydrogenase
KEYWORDS       NADP; Oxidoreductase
SUMMARY        #length 390 #checksum 2790

Query Match      4.1%; Score 103; DB 2; Length 390;
Best Local Similarity 21.7%; Pred. No. 2.16e+00;
Matches 15; Conservative 25; Mismatches 26; Indels 3; Gaps 3;

Db 95 ELEDLSRKYQLIHNYLGSVDIPAPDINTNPOTMAWFL-DEYIKITGE-VDFAVFTGKP 152
QY 245 DLLALGTTNIS-IRFYNASNGYISSQAEFTGAGSWELKEDYVVPENAVGTTIQAQT 303

Db 153 SELGGIGVR 161
QY 304 AQAGQGGM 312

RESULT 12
ENTRY   B55926 #type complete
TITLE   DNA binding protein RFX2 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE     08-Sep-1995 #sequence_revision 08-Sep-1995 #text_change
        10-Sep-1997
ACCESSIONS B55926
REFERENCE   A55926
#authors   Reith, W.; Ucla, C.; Barras, E.; Gaud, A.; Durand, B.;
        Herrero-Sanchez, C.; Kober, M.; Mach, B.
#journal   Mol. Cell. Biol. (1994) 14:1230-1244
#title     RFX1, a transactivator of hepatitis B virus enhancer I,
        belongs to a novel family of homodimeric and heterodimeric
```

```
#cross-references MUID:94119075
#accession     B55926
#status        preliminary
#molecule_type mRNA
#residues      1-723 #label REI
#cross-references GB:X76091; NID:g452389; PID:g452390
SUMMARY        #length 723 #molecular-weight 80003 #checksum 1814

Query Match      4.1%; Score 103; DB 2; Length 723;
Best Local Similarity 29.0%; Pred. No. 2.16e+00;
Matches 18; Conservative 17; Mismatches 25; Indels 2; Gaps 2;

Db 466 VPSTLTQAIENFAKSLGWL-TNAMSDFPOOVITQTKGVV-SAFAPQTLRRYTSNLHQAQ 523
QY 103 IVALTSGKVNFPPEVVSWLRTAGTSAPSDSILSRFDVSYAAFYTSKRAIALEHVKLS 162

Db 524 AR 525
QY 163 NR 164

RESULT 13
ENTRY   S17854 #type complete
TITLE   NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) 75K chain
        precursor - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE     30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
        26-Feb-1999
ACCESSIONS S17854; S16382
REFERENCE   S17854
#authors   Chow, W.; Ragan, I.; Robinson, B. H.
#journal   Eur. J. Biochem. (1991) 201:547-550
#title     Determination of the cDNA sequence for the human
        mitochondrial 75-kDa Fe-S protein of NADH-coenzyme Q
        reductase.
#cross-references MUID:92037608
#accession  S17854
#molecule_type mRNA
#residues  1-727 #label CHO
#cross-references EMBL:X61100; NID:g38078; PID:g38079

GENETICS
#gene       GDB:NDUFS1
#cross-references GDB:132062; OMIM:157655
#map_position 2q33-2q34
CLASSIFICATION #superfamily NADH dehydrogenase (ubiquinone) chain 11
KEYWORDS       iron-sulfur protein; membrane-associated complex;
        metalloprotein; mitochondrion; NAD; oxidoreductase
FEATURE
1-23          #domain transit peptide (mitochondrion) #status
        predicted #label TNP\
24-727        #product NADH dehydrogenase (ubiquinone) 75K chain
        #status predicted #label MAT
SUMMARY        #length 727 #molecular-weight 79573 #checksum 8774

Query Match      4.1%; Score 105; DB 2; Length 727;
Best Local Similarity 28.6%; Pred. No. 1.28e+00;
Matches 26; Conservative 21; Mismatches 39; Indels 5; Gaps 5;

Db 183 IRFASIEIAGVDDILGTTGRGN-DMQVGTYTEKMFMSLSGNIIIDICPVGALT-SKPVAFTA 240
QY 24 VRYKISAGSCPLSTAGPSVVKFQDNPVGSQTSFAGLHLRVFDP-STGALVDSKSTAFST 82

Db 241 RPWETKRTSIDVMDAVGNSIVVSTRT-GEV 270
QY 83 SN-DTTSAAVFSFMSNLTNRRIVAILTSKV 112

RESULT 14
ENTRY   A33552 #type complete
TITLE   NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) 75K chain
        precursor - bovine
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
```

Search completed: Wed Sep 1 16:03:23 1999
Job time : 41 secs.

WILEY

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:03:42 1999; MasPar time 15.64 Seconds
Tabular output not generated. 505,902 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (1-372) from PCTUS9913024.pep (1 of 12)
Perfect Score: 2542
Sequence: 1 MEKFAEFGQGYVQTFPLE.....LPTQASSRTGKVFQGEFREV 372
Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-gensseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 34.285; Variance 172.028; scale 0.199

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	300	11.8	65	19	R97371	Phage T4 ORFX gene pr
2	128	5.0	510	6	R32658	PS2 protein from C.me
3	109	4.3	330	22	W20625	H. pylori cytoplasmic
4	107	4.2	2247	25	W27126	Equine rhinovirus 1 (
5	101	4.0	269	26	W38306	subtilisin 309 (BLS30
6	101	4.0	269	23	W22755	subtilisin 309 (BLS30
7	99	3.9	269	2	R10444	S004 mutant of the su
8	99	3.9	269	2	R10445	S014 mutant of the su
9	99	3.9	269	2	R10446	S020 mutant of the su
10	99	3.9	270	2	R10568	S20 mutant of the sub
11	99	3.9	270	2	R10441	S12 mutant of the sub
12	99	3.9	270	2	R10567	S19 mutant of the sub
13	99	3.9	270	2	R10442	S024 mutant of the su
14	99	3.9	294	17	R88003	Delta-endotoxin MKHK
15	99	3.9	571	29	W55665	H. pylori ORF hp2el09
16	99	3.9	1213	30	W55735	H. pylori ORF 07ee507

17	97	3.8	269	9	R46346	Subtilisin 309 Serine	5.77e+01
18	97	3.8	269	9	R46345	PB92 serine protease;	5.77e+01
19	97	3.8	473	15	R84637	Shc protein.	5.77e+01
20	97	3.8	473	21	W14005	Human SHC protein.	5.77e+01
21	97	3.8	474	17	R37243	SHC phosphotyrosine b	5.77e+01
22	97	3.8	921	25	W22863	Bacillus stearothermo	5.77e+01
23	96	3.8	1693	35	W76368	Hepatitis E virus hol	6.66e+01
24	94	3.7	269	16	R91682	Subtilisin 309 loop 5	8.84e+01
25	93	3.7	269	26	W38314	BLS309 mutant Arg170L	1.02e+02
26	93	3.7	269	23	W22766	BLS309 mutant Ser57Pr	1.02e+02
27	93	3.7	269	23	W22763	BLS309 mutant Arg170L	1.02e+02
28	93	3.7	269	26	W38317	BLS309 mutant Ser57Pr	1.02e+02
29	94	3.7	577	21	W14494	Urease subunit C.	8.84e+01
30	93	3.7	597	2	R12384	Bovine pancreatic cho	1.02e+02
31	94	3.7	664	13	R62486	Human sVEGF-R11.	8.84e+01
32	94	3.7	668	28	W47037	Soluble truncated VEG	8.84e+01
33	94	3.7	764	35	W05679	KDR protein sequence.	8.84e+01
34	94	3.7	788	28	W47038	Soluble truncated VEG	8.84e+01
35	94	3.7	789	13	R62488	Truncated KLD sVEGF-R	8.84e+01
36	94	3.7	848	27	W38240	VEGF receptor KDR bin	8.84e+01
37	94	3.7	1216	21	W14497	Urease protein.	8.84e+01
38	94	3.7	1358	5	R26999	Novel type III Rtk en	8.84e+01
39	94	3.7	1356	32	W59275	Human KDR protein.	8.84e+01
40	91	3.6	269	26	W38310	BLS309 mutant Ser57Pr	1.35e+02
41	91	3.6	474	27	W39089	Human ShcC mutant A13	1.35e+02
42	92	3.6	520	1	P94617	Neutral protease enco	1.17e+02
43	92	3.6	521	3	P51009	Sequence of neutral p	1.17e+02
44	92	3.6	543	26	W25739	Soybean protoporphyr	1.17e+02
45	91	3.6	1810	16	R94563	Chicken cytotactin.	1.35e+02

ALIGNMENTS

RESULT 1
ID R97371 standard; Protein; 65 AA.
AC R97371;
DT 07-JAN-1997 (first entry)
DE Phage T4 ORFX gene product. gp34.
KW Phage T4; tail fibre protein; nanotechnology; nano-structure;
filter; molecular sieve.
OS Bacteriophage T4.
PN M09611947-A1.
PD 25-APR-1996.
PR 13-OCT-1995; U13023.
PF 13-OCT-1994; US-322760.
PA (GOLD/) GOLDBERG E B.
PI Goldberg EB;
DR N-PSDB; T29053.
PT New proteins derived from T4 phage tail fibre proteins - that can
self assemble into nano-structure(s), useful as filters etc, also
corresponding DNA
PS Claim 7; Fig 7; 83pp; English.
CC A protein (R97371) of unspecified function is the product of open
reading frame X of the tail fibre protein gene region (see also
CC T29053) of phage T4. This gene region also includes open reading
CC frames for tail fibre proteins (see also R97370 and R97372-74).
CC Tail fibre proteins (native or modified) can be produced in large
CC quantities in microbial cells and used as building blocks of strong,
CC stable nanostructures.
SQ Sequence 65 AA;

Query Match 11.88; Score 300; DB 19; Length 65;
Best Local Similarity 82.18; Pred. No. 8.92e-14;
Matches 46; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Db 1 mekfmaefgqgyvqtfple.....lptqassrtgkvgfgeqfrev 56
QY 1 MEKFAEFGQGYVQTFPLE.....LPTQASSRTGKVFQGEFREV 56

RESULT 2
ID R32658 standard; Protein; 510 AA.

AC R32658;
DT 18-JUN-1993 (first entry)
DE PS2 protein from C.melasscola ATCC 17965.
KW NADPH dependent glutamate dehydrogenase; Corynebacterium glutamicum;
KW cs2 gene; secretory peptide.
OS Corynebacterium melasscola ATCC 17965.
FH Key Location/Qualifiers
FT Peptide 1..30
FT /label= signal
FT /note= "for protein secretion"
FT cleavage_site 26..30
FT /note= "putative"
FT protein 31..510
FT /label= PS2
FT
FT
PN W09303158-A.
PD 18-FEB-1993.
PF 29-JUL-1992; F00744.
PR 30-JUL-1991; FR-009652.
PR 02-AUG-1991; FR-009870.
PA (ORSA-) ORSAN.
PI Duchiron F, Guyonvarch A, Joliff G, Leblon G, Relano P;
FI Renaud M;
DR WPI; 93-076515/09.
DR N-PSDB: Q37239.
PT Protein expression and secretion system in corynebacteria -
PT transformed with cassette contg. signal sequence of protein PS1
PT or 2 in chromosome or plasmid DNA region, for repetitive
PT synthetic aminoacid polymer and polypeptide prodn.
PS Claim 1; Fig 12; 95pp; French.
CC N-terminal signal peptide of PS2 is useful in recombinant protein
CC production. It results in secretion of the protein from the
CC transformant. Since PS2 is associated with the bacterial cell wall,
CC a portion of the expressed product will remain attached to the wall
CC and can be recovered (without lysis) by treatment with a surfactant.
SQ Sequence 510 AA;

Query Match 5.0%; Score 128; DB 6; Length 510;
Best Local Similarity 23.3%; Pred. No. 5.54e-01;
Matches 52; Conservative 55; Mismatches 100; Indels 16; Gaps 13;

Db 134 esvayalkvdeat-aafeyarnalrdaai-sinpdgsin-pdtsinllidaanaanrtid 190
Qy 74 DSKSYAFSTSDTSAFVFMNSLTNNRIVAILTSGRVNFPPEVWSLRTAGTSAPPSD 133
Db 191 r-aeie-dvahltyqt--dialetpqlayafqdlkalqaevdadfewlgefgidq-edg 244
Qy 134 SILSKFDVSYAFTSSRAALEHVKLSNR-KSTDDTQTILDVVFDSLEDVGGATGPPRG 192
Db 245 ny--vqryhlpavealkaedvdaavaieplradslaknleaqksdvlvrlqlerataqr 302
Qy 193 TYESVEQF-MSAVGNTNDEI-ARLPTSAISKLS-DYNLIPGDVLYLKAQLYADADLLAL 249
Db 303 dclrvaeifstaryvelyenvennvntklrhysaallpn 345
Qy 250 GTTWISIRFYNASNGYISSTQAEFTGQAGSWELKEDY-VVYVE 291

RESULT 3
ID W20625 standard; Protein: 330 AA.
AC W20625;
DE 14-JUL-1997 (first entry)
DE H. pylori cytoplasmic protein, O2ae11612orf26.
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacterium; life cycle; activator;
KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
KW diagnosis.
OS Helicobacter pylori.
PN W09640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996; U09122.
PR 07-JUN-1995; US-487032.
PR 01-APR-1996; US-630405.
PA (ASTR) ASTRA AB.

PI Berglindh OT, Smith D, Mellgaard BL;
DR WPI; 97-052306/05.
DR N-PSDB: T67878.
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 61; Page 1051-52; 1481pp; English.
CC This sequence represents a H. pylori cytoplasmic protein involved in
CC cofactor metabolism.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 330 AA;

Query Match 4.3%; Score 109; DB 22; Length 330;
Best Local Similarity 28.0%; Pred. No. 1.00e+01;
Matches 23; Conservative 26; Mismatches 28; Indels 5; Gaps 5;

Db 164 vdllaplanm-mdgnvls-lrkalkdkgfyhtplms-ystkfassygpfdrdvanspsfg 220
Qy 104 VAILTSGRVNFPPEVWSLRTA-CTSAPPSDILSRFDVSYAFTSSKRAIALEHVKLS 162
Db 221 drksymdyanqkeallesled 242
Qy 163 NRKSTD-DYQITILDVVFDSLED 183

RESULT 4
ID W27126 standard; Protein: 2247 AA.
AC W27126;
DT 11-FEB-1998 (first entry)
DE Equine rhinovirus 1 (ERHV1), polypeptide amino acid sequence.
KW Equine rhinovirus 1; ERHV1; foot-and-mouth disease virus; vaccine; horse;
KW diagnosis; antigens; polypeptide; enzyme-linked immunosorbent assay;
KW recombinant protein.
OS Equine rhinovirus 1.
FH Key Location/Qualifiers
FT Cleavage_site 207..208
FT Cleavage_site 287..288
FT Cleavage_site 517..518
FT Cleavage_site 743..744
FT Cleavage_site 991..992
FT Cleavage_site 1007..1008
FT Cleavage_site 1143..1144
FT Cleavage_site 1158..1159
FT Cleavage_site 1552..1553
FT Cleavage_site 1577..1578
FT Cleavage_site 1782..1783
FT Peptide 1..207
FT /label= L
FT Peptide 208..287
FT /label= VP4
FT Peptide 288..517
FT /label= VP2
FT Peptide 518..743
FT /label= VP3
FT Peptide 744..991
FT /label= VP1
FT Peptide 992..1007
FT /label= 2A
FT Peptide 1008..1143
FT /label= 2B
FT Peptide 1144..1458

FT Peptide /label= 2C
 FT 1459..1552
 FT /label= 3A
 FT 1553..1577
 FT /label= 3B
 FT 1578..1782
 FT /label= 3C
 FT 1783..2246
 FT /label= 3D
 PN WQ9722701-A1.
 PD 26-JUN-1997.
 PF 18-DEC-1996: AU0815.
 PR 18-DEC-1995: AU-007201.
 PA (UYME) UNIV MELBOURNE.
 PI Crabb BS, Feng L, Studdert MJ;
 DR WPI: 97-341692/31.
 DR N-PSDB: T83178.
 DR Genomic sequence of equine rhinovirus 1 - and derived proteins or
 PT virus-like particles, useful in vaccines and as diagnostic agents
 PS Claim 2: Pages 30-32; 60pp; English.
 CC The present sequence represents the amino acid sequence of the
 CC polypeptide of equine rhinovirus 1 (ERHV1). The taxonomic status
 CC of ERHV1 is unclear, as physicochemical studies have shown that the
 CC nucleic acid density and base composition of ERHV1 differs from other
 CC rhinoviruses. To this end, the nucleotide sequence encoding the
 CC polypeptide of ERHV1 was deduced. Analysis of this sequence suggests that
 CC ERHV1 is more closely related to foot-and-mouth disease virus. Individual
 CC ERHV1 proteins can be used to make vaccines to protect horses (and
 CC possibly other animals) against ERHV1. Oligonucleotide primers and probes
 CC can be used for diagnosis of ERHV1 or related viruses, while antigens of
 CC the ERHV1 polypeptide can be used to detect ERHV1-specific antibodies in
 CC the blood, particularly in enzyme-linked immunosorbent assay. They can
 CC differentiate between infected animals and those vaccinated with ERHV1
 CC vaccines (the infected animals will have antibodies reactive with
 CC non-capsid proteins but vaccinated animals will not). Fragments of the
 CC DNA sequence represent individual genes of the virus and can be
 CC expressed in host systems to produce recombinant proteins. Virus like
 CC particles containing the individual ERHV1 proteins, can also be used as
 CC vectors for delivering therapeutic or other useful agents, including
 CC vaccinating epitopes from other pathogens or reproductive hormones.
 SQ Sequence 2247 AA;
 Query Match 4.2%; Score 107; DB 25; Length 2247;
 Best Local Similarity 23.4%; Pred. No. 1.35e+01;
 Matches 22; Conservative 33; Mismatches 32; Indels 7; Gaps 7;
 Db 1041 tpspfykamkflatlavaamrtkdppvvmliaf-gle-vfdtgffsfyqeklqp 1098
 QY 142 SYAAFTSSKRAIALEHVKLSNRKSTDDYQTILDVY-FDSLEDVGATGFPRTY-ESVEQ 199
 Db 1099 ymktpgkisdlytdaataaaiqp-gvysfvs 1131
 QY 200 FMSAVGVTNDE-IARLPTSAA-ISKLSIDYNLIPG 231
 RESULT 5
 ID W38306 standard; protein: 269 AA.
 AC W38306;
 DT 27-MAR-1998 (first entry)
 DE Subtilisin 309 (BLS309) mutant Arg170Phe (in BASBPN numbering).
 KW Variant; mutant; subtilase; enzyme; BASBPN; storage stability;
 KW wash performance; detergent; hydrophobic domain; subtilisin 309;
 BLS309.
 OS Bacillus lentus.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc_difference 164
 FT /note= "wild type Arg replaced by Phe"
 PN W09634935-A2.
 PD 07-NOV-1996.
 PF 12-APR-1996: E01610.
 PR 05-MAY-1995: EP-201161.
 PA (UNIL) UNILEVER NV.

PA (UNIL) UNILEVER PLC.
 PI Bauditz P, Markvardsen P, Sierkstra LN, Von Der Osten C,
 DR Klugkist J;
 DR WPI: 96-506142/50.
 PT Detergent compositions comprising Subtilisin variants - which
 PT exhibit improved stability and/or wash performance
 PS Claim 1: Page -: 91pp; English.
 CC The present subtilase BLS309 variant, which has improved storage
 CC stability and/or improved performance in detergents, has a residue
 CC in the vicinity a wild type hydrophobic domain substituted with a
 CC more hydrophobic residue. The hydrophobic domain comprises the
 CC residues Pro129, Pro131, Ile165, Tyr167, Tyr171 of BLS309 (in BASBPN
 CC numbering). Detergent compositions containing the mutant enzyme
 CC have improved storage stability, while retaining or improving their
 CC wash performance. This variant has a wash performance improvement
 CC factor of 3.3 compared to wild type BLS309.
 CC N.B. Sequence not given in the specification, but constructed using
 CC the wild type BLS309 sequence, Genbank A08332.
 SQ Sequence 269 AA;
 Query Match 4.0%; Score 101; DB 26; Length 269;
 Best Local Similarity 32.8%; Pred. No. 3.25e+01;
 Matches 19; Conservative 21; Mismatches 14; Indels 4; Gaps 4;
 Db 159 isypafyanamavgatdq-n-nnrasfsgyagldivapgv-nvqst-ypgstvasln 212
 QY 141 VSYAAFTSSKRAIALEHVKLSNRKSTDDYQTILDVY-FDSLEDVGATGFPRTY-ESVE 198
 RESULT 6
 ID W22755 standard; protein: 269 AA.
 AC W22755;
 DT 11-SEP-1997 (first entry)
 DE Subtilisin 309 (BLS309) mutant Arg170Phe (in BASBPN numbering).
 KW Variant; mutant; subtilase; enzyme; BASBPN; storage stability;
 KW wash performance; detergent; hydrophobic domain; subtilisin 309;
 BLS309.
 OS Bacillus lentus.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Key misc_difference 164
 FT /note= "wild type Arg replaced by Phe"
 PN W09634946-A1.
 PD 07-NOV-1996.
 PF 02-MAY-1996: DK0207.
 PR 05-MAY-1995: DK-000519.
 PR 12-APR-1996: DK-000421.
 PA (NOVO) NOVO-NORDISK AS.
 PI Bauditz P, Klugkist J, Markvardsen P, Sierkstra LN;
 PI Von Der Osten C;
 DR WPI: 96-506152/50.
 PT A subtilase variant with amino acid substitutions in or near
 PT hydrophobic domain - has improved storage stability and/or improved
 PT performance in detergents, whilst retaining or improving its wash
 PT performance
 PS Claim 1: Page -: 91pp; English.
 CC The present subtilase BLS309 variant, which has improved storage
 CC stability and/or improved performance in detergents, has a residue
 CC in the vicinity a wild type hydrophobic domain substituted with a
 CC more hydrophobic residue. The hydrophobic domain comprises the
 CC residues Pro129, Pro131, Ile165, Tyr167, Tyr171 of BLS309 (in BASBPN
 CC numbering). Detergent compositions containing the mutant enzyme
 CC have improved storage stability, while retaining or improving their
 CC wash performance. This variant has a wash performance improvement
 CC factor of 3.3 compared to wild type BLS309.
 CC N.B. Sequence not given in the specification, but constructed using
 CC the wild type BLS309 sequence, Genbank A08332.
 SQ Sequence 269 AA;
 Query Match 4.0%; Score 101; DB 23; Length 269;
 Best Local Similarity 32.8%; Pred. No. 3.25e+01;
 Matches 19; Conservative 21; Mismatches 14; Indels 4; Gaps 4;

thuringiensis (B.t.). This sequence represents the MKHHK delta-endotoxin, and has a molecular weight of 33 kDa. Delta-endotoxins have molecular weights ranging from 27-140 kDa, and kill insect larvae upon ingestion. These sequences can be used in insecticidal compositions, which can be used to control lepidopteran and coleopterian pests. These compositions can optionally contain spores from a biologically pure B.t. strain. The compositions also contain at least two other delta-endotoxins, each having a molecular weight of 130 kDa. The delta-endotoxins increase the spectrum of available biopesticides, particularly to any given insect pest.

SQ Sequence 294 AA;

Query Match 3.9%; Score 99; DB 17; Length 294;

Best Local Similarity 26.0%; Pred. No. 4.33e+01;

Matches 25; Conservative 28; Mismatches 38; Indels 5; Gaps 5;

Db 119 tqsltvskfsfnfvagaeannisfsvgeqnlstektteststlmripqpqvsvrpt 178

Qy 69 TQAL-VDSK-SYAFSTNDTTSAAF-VSFMSLTNNRIVAILTSGKVNFPPEVWS-WLRT 124

Db 179 akrveislfaiprigneisgfvgtlptltsnsh 214

Qy 125 AGTSAFP-SDSILSRFDVSYAAFTSSKRALEHV 159

RESULT 15

ID W55665 standard; Protein: 571 AA.

AC W55665;

DT 24-JUN-1998 (first entry)

DE H. pylori ORF hpzel0911_4882027_c2_87 secreted protein.

KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;

KW identification; binding compound; bacteria; life cycle; activator;

KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.

OS Helicobacter pylori.

PN W09737044-A1.

PD 09-OCT-1997.

PF 27-MAR-1997; U052223.

PR 06-DEC-1996; US-761318.

PR 29-MAR-1996; US-625811.

PR 02-APR-1996; US-758731.

PR 25-OCT-1996; US-736905.

PR 28-OCT-1996; US-738859.

PA (ASTR) ASTRA AB.

PI Alm RA, Smith D;

DR WPI; 97-503122/46.

DR N-PSDB: V25074.

PT Helicobacter pylori nucleic acid sequences and encoded

PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori

PT infection and for diagnosis of H. pylori infection

PS disclosure: Pages 871-872; 1145pp; English.

CC This is the sequence of a Helicobacter pylori secreted protein.

CC The protein may be used in a vaccine to prevent or treat H. pylori

CC infection or to identify H. pylori polypeptide binding compounds,

CC useful as potential H. pylori life cycle activators or inhibitors. The

CC DNA and probes derived from it may be used for the identification of

CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic

CC acid sequences complementary to the DNA act as antisense sequences and

CC can be used to prevent the translation of H. pylori mRNA. Antibodies

CC against the protein can be used in immunoassays to evaluate the abundance

CC and distribution of H. pylori-specific antigens. The genomic sequence of

CC H. pylori (ATCC 55679) was determined from overlapping contigs generated

CC by mechanically shearing the bacterial DNA. The sequences were analysed

CC for ORF of at least 180 nucleotides, and the predicted coding regions

CC defined by computer evaluation. To identify likely H. pylori antigens for

CC vaccine development, the amino acid sequences predicted from various ORF

CC were analysed for significant homology to other known or exported

CC membrane proteins. Having identified and determined the sequences of

CC interest, particular regions can be isolated from H. pylori by PCR

CC amplification for recombinant polypeptide production, e.g. in E. coli

CC hosts.

SQ Sequence 571 AA;

Query Match 3.9%; Score 99; DB 29; Length 571;

Best Local Similarity 21.6%; Pred. No. 4.33e+01;
Matches 30; Conservative 43; Mismatches 58; Indels 8; Gaps 7;
Db 190 lnfnnnqfngsfeigakntifn-nan-fnnstsfnnssatts--fvgdftnansnl 245
Qy 44 VKF-QDNVPVGSQTFSAGLHLRFVDPSTGALVDSKSYAFSTNDTTSAAAFVS-FMNSLTNN 101
Db 246 qiagnavfgnsgntqntanfnntgsvniagnatfdnv-vfnspntsvkgkvtlnnltl 304
Qy 102 RIVAILTSGKVNFPPEVVSWLRTAGTSAPPSDSILSRFDVSYAAFTSSKRAIALEHVKL 161
Db 305 knlnaplsfgdg-tivfsa 322
Qy 162 SNRKSTDDYQITLDVWFDS 180

Search completed: Wed Sep 1 16:04:42 1999

Job time : 60 secs.

W P S R E H (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Sep 1 16:01:19 1999; MasPar time 25.10 Seconds
Tabular output not generated. 808.836 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (1-372) from PCTUS9913024.pep (1 of 12)
Perfect Score: 2942
Sequence: 1 MEKFAEFGGQYVTFPFLSE.....LPTQASSKTGKVGQEFREV 372

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sprenb19
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phase 10:sp-plant 11:sp-rodent 12:sp_unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 47.705; Variance 93.090; scale 0.512

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	397	15.6	58	9 Q38400	TAIL FIBER GENE 35 (FR 1.12e-52)	
2	391	15.4	58	9 Q38425	TAIL FIBER GENE 35 (FR 1.70e-51)	
3	382	15.0	58	9 Q38393	TAIL FIBER GENE 35 (FR 9.94e-50)	
4	128	5.0	510	2 Q04985	PS2 PROTEIN PRECURSOR. 3.35e-04	
5	111	4.4	373	5 Q61942	T24A6.11 PROTEIN. 7.82e-02	
6	113	4.4	227	14 Q66775	POLYPROTEIN. 4.22e-02	
7	113	4.4	2248	14 Q66774	POLYPROTEIN. 4.22e-02	
8	113	4.4	2497	14 Q88578	NONSTRUCTURAL POLYPROT 1.44e-01	
9	109	4.3	1007	4 Q43424	GLUTAMATE RECEPTOR DEL 2.63e-01	
10	107	4.2	920	2 Q45664	REVERSE TRANSCRIPTASE 2.63e-01	
11	106	4.2	1300	11 Q08906	POLYPROTEIN. 3.55e-01	
12	107	4.2	2247	14 Q39818	NONSTRUCTURAL POLYPROT 3.55e-01	
13	106	4.2	2493	14 Q39818	NONSTRUCTURAL POLYPROT 3.55e-01	
14	106	4.2	2493	14 Q66594	NONSTRUCTURAL POLYPROT 3.55e-01	
15	106	4.2	2493	14 Q66592	PUTATIVE NONSTRUCTURAL 8.59e-01	
16	103	4.1	277	1 Q58196	277AA LONG HYPOTHETICA 8.59e-01	
17	103	4.1	325	2 Q88073	PUTATIVE TRANSCRIPTION 4.78e-01	
18	105	4.1	1546	2 Q45996	SCAFFOLDING PROTEIN PR 1.53e+00	
19	101	4.0	548	10 Q39000	AMIDOPHOSPHORIBOSYLTRA 1.53e+00	
20	101	4.0	997	5 Q44825	H17B01.4 PROTEIN. 1.53e+00	

21 101 4.0 1004 2 084646 EXODEOXYRIBONUCLEASE V 1.53e+00
22 101 4.0 1577 2 054178 GLUCOSYLTRANSFERASE. 1.53e+00
23 98 3.9 203 10 049136 CELL DIVISION PROTEIN 3.60e+00
24 98 3.9 311 2 052041 CATECHOL 1,2-DIOXYGENA 3.60e+00
25 98 3.9 321 5 077113 PERIOD (FRAGMENT). 3.60e+00
26 98 3.9 383 5 016306 T21H3.5 PROTEIN. 3.60e+00
27 100 3.9 466 11 061787 ORF 2. 2.04e+00
28 98 3.9 572 2 073543 HYPOTHETICAL 62.9 KD P 3.60e+00
29 100 3.9 823 10 039594 FUS1 PROTEIN. 2.04e+00
30 100 3.9 863 4 014689 KIAA0184 PROTEIN (FRAG 2.04e+00
31 99 3.9 1007 11 061825 GLUTAMATE RECEPTOR CHA 2.72e+00
32 99 3.9 1007 11 062841 GLUTAMATE RECEPTOR DEL 2.72e+00
33 99 3.9 1008 11 063226 GLUTAMATE RECEPTOR SUB 2.72e+00
34 100 3.9 1281 11 088915 L1 RETROTRANSPOSON, TF 2.04e+00
35 99 3.9 1281 11 088914 L1 RETROTRANSPOSON, TF 2.72e+00
36 99 3.9 1281 11 054850 ENDONUCLEASE/REVERSE T 2.72e+00
37 99 3.9 1281 11 088913 L1 RETROTRANSPOSON, TF 2.72e+00
38 100 3.9 1300 11 061785 LIMD-A13 REPETITIVE SE 2.04e+00
39 100 3.9 1693 14 081876 COMPLETE GENOME SEQUE 2.04e+00
40 98 3.9 2567 13 090574 FILAMIN. 3.60e+00
41 97 3.8 311 2 051960 CATECHOL OXYGENASE. 4.76e+00
42 97 3.8 413 2 048595 PLASMOD. PTR2030 LLAI M 4.76e+00
43 97 3.8 476 4 075674 TOM1-LIKE PROTEIN. 4.76e+00
44 97 3.8 583 4 015290 P66SHC. 4.76e+00
45 97 3.8 1350 5 015884 INVARIANT SURFACE GLYC 4.76e+00

ALIGNMENTS

RESULT 1
ID Q38400 PRELIMINARY; PRT: 58 AA.
AC Q38400;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE TAIL FIBER GENE 35 (FRAGMENT).
OS BACTERIOPHAGE OX2.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85215500.
RA RIEDE I., DREXLER K., ESCHBACH M.L.;
RT "The nucleotide sequences of the tail fiber gene 36 of bacteriophage T2 and of genes 36 of the T-even type Escherichia coli phages K3 and OX2".
RL NUCLEIC ACIDS RES. 13:605-616(1985).
DR EMBL; X01753; E7867; -.
FT NON_TER 1
SQ SEQUENCE 58 AA; 6206 MW; 29127C56 CRC32;

Query Match 15.6%; Score 397; DB 9; Length 58;
Best Local Similarity 98.3%; Pred. No. 1.12e-52;
Matches 57; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
DB 1 SFSEVSRNGGSKPAEFGVGNVYCVESASPPDMLVLPQTQASSKTGKVGQEFREV 58
QY 315 SFSEVSRNGGSKPAEFGVGNVYCVESASPPDMLVLPQTQASSKTGKVGQEFREV 372

RESULT 2
ID Q38425 PRELIMINARY; PRT: 58 AA.
AC Q38425;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE TAIL FIBER GENE 35 (FRAGMENT).
OS BACTERIOPHAGE T2.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85215500.

RIEDEL I., DREXLER K., ESCHBACH M. L.,
RT THE NUCLEOTIDE SEQUENCES OF THE TAIL
RT T2 AND OF GENES 36 OF THE T-EVEN TYPE
RT OX2.
RT ESCHERICHIA COLI PHAGES K3 AND
RT
RL NUCLEIC ACIDS RES. 13:605-616(1985).
DR EMBL: X01755; G15190. --
DR
FT NON_TER 1 1
SQ SEQUENCE 58 AA; 6236 MW; E15E16A4 CRC32;

```
Query Match      15.4%; Score 391; DB 9; Length 58;
Best Local Similarity 96.6%; Pred. No. 1.70e-51;
Matches 56; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```

Db      1 SFSEVSRNGISISKPAEFGVNGIRVNVCESASPPDIMVLPTQASSKTGKVGFGEFREV 58
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy     315 SFSEVSRNGGISISKPAEFGVNGIRVNVICESASPPDIMVLPTQASSKTGKVGFGEFREV 372

```

RESULT		3		
ID	Q38993		PRELIMINARY:	PRT; 58 AA.
AC	Q38993;			
DT	01-NOV-1996 (TREMREL. 01, CREATED)			
DT	01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)			
DE	TAIL FIBER GENE 35 (FRAGMENT).			
OS	BACTERIOPHAGE K3.			
OC	VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE.			
RN	[1]			
RP	SEQUENCE FROM N.A.			

RA MEDLINE; 63213500.
RA RIEDE I., DREXLER K., ESCHBACH M.L.;
RT "The nucleotide sequences of the tail fiber gene 36 of bacteriophage
RT T2 and of genes 36 of the T-even type *Escherichia coli* phages K3 and
RT OX2.";
RL NUCLEIC ACIDS RES. 13:605-616(1985).

PR ORDER, 201754, 27055, 1
 FT NON_TER 1
 SQ SEQUENCE 58 AA; 6246 MW; 8F8EBCDA CRC32;
 Query Match 15.08; Score 382; DB 9; Length 58;
 Best Local Similarity 94.8%; pred. No. 9.94e-50;
 Matches 55; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

Db      1 SFSEVSRNGSISKPAEFGINGIRVNYICESASPPDIIVLPTQASSKTGKVFQGEFREV 58
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy     315 SFSEVSRNGGISKPAEFGVNGIRVNYICESASPPDIMVLPTQASSKTGKVFQGEFREV 372

```

RESULT	4	PRELIMINARY;	PRT;	510 AA.
ID	Q04985			
AC	Q04985;			
DT	01-NOV-1996	(TREMREL. 01, CREATED)		
DT	01-NOV-1996	(TREMREL. 01, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998	(TREMREL. 08, LAST ANNOTATION UPDATE)		
DE	PS2 PROTEIN PRECURSOR.			
OS	CSP2.			
OS	CORYNEBACTERIUM GLUTAMICUM.			
OC	BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIIDAE;			
OC	ACTINOMYCETALES; CORYNEBACTERINEAE; CORYNEBACTERIACEAE;			
OC	CORYNEBACTERIUM.			

SEQUENCE FROM N.A.
STRAIN=ATCC17965.
RX MEDLINE: 94018618.
RT PEYRET J.L., JOLIFF G., MATHIEU L., SHECHTER E., LEBLON G.;
RT "Characterization of the csbB gene encoding PS2, an ordered
RT surface-layer protein in *Cornebacterium glutamicum*.";
RL MOL. MICROBIOL. 9:97-109(1993).

Accession	Protein	Length	Signal	Potential
EMBL; X09103; G40469; .	PS2 PROTEIN.	31	510	
DR	SEQUENCE	510 AA	55425 MW	146DDF70 CRC32;
KW	SIGNAL.	1	30	POTENTIAL.
FT	SIGNAL	31	510	
FT	CHAIN			
SQ	SEQUENCE	510 AA	55425 MW	146DDF70 CRC32;

Query Match 5.0%; Score 128; DB 2; Length 510;
Best Local Similarity 23.3%; Pred. No. 3.35e-04;
Matches 52; Conservative 55; Mismatches 100; Indels 16; Gaps 13;

```

Db      134  ESVAYALKVDQEAT-AFEYARNALPDAAI-SINPDGSIN-PDTSINLILIDAANANRTD 190
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qv      74  DKSYAFESTNDTSAAFVFEMNSITNRIVAILTSGKVNFPPEWSITRAGTSAFPSD 133

```

191 R--AEIE-DYAHLYTQT--DIALETPQLAYAFQDLKALQAEVDADFEWLGEFGIDQ-EDG 244
db

QY 134 SILSRFDVSAAFYTSSKRAIALEHVKLSNR-KSTDDYQTI LDWFDSDLEDVGATGFPGRG 192

Db 245 NY--VQRTHPAVEALKAEVDARVAAIEPLRADSIKNLEAQKSDVLVRQLFLERATAQR 302

QY 193 TYESVEQF-MSAVGGTND E I-ARLPTSA AISKLS-DYNLIPGDVLYLKAQLYADADLLAL 249

Db 303 DTLRVEATFSTARYVELYENVENVNENKTLRQHYSALIPN 345

RESULT 5
ID 061942
PRELIMINARY;
PRT; 373 AA.

AC	001942;
DT	01-AUG-1998 (TREMREL. 07, CREATED)
DT	01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
DT	01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE	T24A6.11 PROTEIN.

GN 12486.11.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELOPERINAE; CAENORHABDITIS.

RP [1] SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;
BY WERTZ INC. 041507310

RA WILSON R., AINSCOU
RA BONEFIELD T, PUERTO

RA BONFIELD J., BURTO
RA CRAXTON M., DEAR S

RA GARDNER A., GREEN
RA JONES M., KERSHAW

RA LIGHTNING J., LLOYD

RA SMALDON N., SMITH
RA THIERRY-MIEC 1 E

RA THIERRY-MIEG J., T
RA WATSON A., WEINSTO

RT "2.2 Mb of contiguous
RT elegans.";

RL NATURE 368:32-38(1

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;
RA TAICH A.; GATTING

RL
RN
SUBMITTED (JUN-1999
[3]

RP SEQUENCE FROM N.A.

RA WATERSTON R.;
PI SUBMITTED / MAY-100

RL SUBMITTED (MAY-1999
DR EMBL: AF068713: G3

SQ SEQUENCE 3/3 AA;

Query Match

Matches 36; **Conse**

Db 157 SPEV-TRVSSTST

Qy 115 PPEVSWLRTAGT
:||| :||| :|||

[illegible]

—
—
..
—
..
..

..
..

..

Qy 170 YOTLDVVDSLEDV-GATG-FP-RGTVESVEQFMSAVGGTNDTEIARLPTSAAIKLSYD 226

Db 275 EITDEEFLVTIVILFCDPAISGLSHALS 304

Qy 227 NLIQCDVLYLKAQLYADADLLALGTTNISI 256

RESULT 6

ID Q66775 PRELIMINARY; PRT; 2227 AA.

AC Q66775;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE POLYPROTEIN.

OS EQUINE RHINOVIRUS TYPE 1.

OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;

OC APHTHOVIRUS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-PERV;

RX MEDLINE; 96332508.

RA WUTZ G., AUER H., NOWOTNY N., SKERN T., KUECHLER E.;

RT "Equine rhinovirus serotypes 1 and 2: relationship to each other and to aphthoviruses and cardioviruses.";

RL J. GEN. VIROL. 77:1719-1730(1996).

DR EMBL; X96870; E235486; -

DR PFAM; PF00073; rhv; 2.

DR PFAM; PF00680; RNA_dep_RNA_pol; 1.

DR PFAM; PF00910; RNA_helicase; 1.

KW POLYPROTEIN.

FT CHAIN 189 268 VP4.

FT CHAIN 269 498 VP2.

FT CHAIN 499 724 VP3.

FT CHAIN 725 972 VP1.

FT CHAIN 973 988 2A.

FT CHAIN 989 1124 2B.

FT CHAIN 1125 1439 2C.

FT CHAIN 1440 1534 3A.

FT CHAIN 1535 1558 3B (VPG).

FT CHAIN 1559 1763 3C (PROTEINASE).

FT CHAIN 1764 2227 3D (RNA POLYMERASE).

SQ SEQUENCE 2227 AA; 247166 MW; A4B1C493 CRC32;

Query Match 4.4%; Score 113; DB 14; Length 2227;

Best Local Similarity 24.5%; Pred. No. 4.22e-02;

Matches 23; Conservative 33; Mismatches 31; Indels 7; Gaps 7;

Db 1022 TSPFYKTAKMLFKLATLAVTAMTKDPVVVVVLIADF-GLE-VFDTGFFSFYQEKLOP 1079

Qy 142 SYAFTYSSKRAIALEHVKLSNRKSTDDYQITLDVV-FDSLEDVGATGFPRTY-ESVEQ 199

Db 1080 YMKTIPIKSELVTDATAAAQIPK-GVYSFVSS 1112

Qy 200 FMSAVGCTNDE-IARLPTSAA-ISKLSYDYNLIPG 231

RESULT 7

ID Q66774 PRELIMINARY; PRT; 2248 AA.

AC Q66774;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE POLYPROTEIN.

OS EQUINE RHINOVIRUS TYPE 1.

OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;

OC APHTHOVIRUS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-PERV;

RX MEDLINE; 96332508.

RA WUTZ G., AUER H., NOWOTNY N., SKERN T., KUECHLER E.;

RT "Equine rhinovirus serotypes 1 and 2: relationship to each other and to aphthoviruses and cardioviruses.";

RL J. GEN. VIROL. 77:1719-1730(1996).

DR EMBL; X96870; E235486; -

DR PFAM; PF00073; rhv; 2.

DR PFAM; PF00680; RNA_dep_RNA_pol; 1.

DR PFAM; PF00910; RNA_helicase; 1.

KW POLYPROTEIN.

FT CHAIN 189 268 VP4.

FT CHAIN 269 498 VP2.

FT CHAIN 499 724 VP3.

FT CHAIN 725 972 VP1.

FT CHAIN 973 988 2A.

FT CHAIN 989 1124 2B.

FT CHAIN 1125 1439 2C.

FT CHAIN 1440 1534 3A.

FT CHAIN 1535 1558 3B (VPG).

FT CHAIN 1559 1763 3C (PROTEINASE).

FT CHAIN 1764 2227 3D (RNA POLYMERASE).

SQ SEQUENCE 2227 AA; 247166 MW; A4B1C493 CRC32;

Query Match 4.4%; Score 113; DB 14; Length 2227;

Best Local Similarity 24.5%; Pred. No. 4.22e-02;

Matches 23; Conservative 33; Mismatches 31; Indels 7; Gaps 7;

Db 1022 TSPFYKTAKMLFKLATLAVTAMTKDPVVVVVLIADF-GLE-VFDTGFFSFYQEKLOP 1079

Qy 142 SYAFTYSSKRAIALEHVKLSNRKSTDDYQITLDVV-FDSLEDVGATGFPRTY-ESVEQ 199

Db 1080 YMKTIPIKSELVTDATAAAQIPK-GVYSFVSS 1112

Qy 200 FMSAVGCTNDE-IARLPTSAA-ISKLSYDYNLIPG 231

RESULT 7

ID Q66774 PRELIMINARY; PRT; 2248 AA.

AC Q66774;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE POLYPROTEIN.

OS EQUINE RHINOVIRUS TYPE 1.

OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;

OC APHTHOVIRUS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-PERV;

RX MEDLINE; 96332508.

RA WUTZ G., AUER H., NOWOTNY N., SKERN T., KUECHLER E.;

RT "Equine rhinovirus serotypes 1 and 2: relationship to each other and to aphthoviruses and cardioviruses.";

RL J. GEN. VIROL. 77:1719-1730(1996).

DR EMBL; X96870; E235486; -

DR PFAM; PF00073; rhv; 2.

DR PFAM; PF00680; RNA_dep_RNA_pol; 1.

DR PFAM; PF00910; RNA_helicase; 1.

KW POLYPROTEIN.

FT CHAIN 189 268 VP4.

FT CHAIN 269 498 VP2.

FT CHAIN 499 724 VP3.

FT CHAIN 725 972 VP1.

FT CHAIN 973 988 2A.

FT CHAIN 989 1124 2B.

FT CHAIN 1125 1439 2C.

FT CHAIN 1440 1534 3A.

FT CHAIN 1535 1558 3B (VPG).

FT CHAIN 1559 1763 3C (PROTEINASE).

FT CHAIN 1764 2227 3D (RNA POLYMERASE).

SQ SEQUENCE 2227 AA; 247166 MW; A4B1C493 CRC32;

Query Match 4.4%; Score 113; DB 14; Length 2227;

Best Local Similarity 24.5%; Pred. No. 4.22e-02;

Matches 23; Conservative 33; Mismatches 31; Indels 7; Gaps 7;

Db 1022 TSPFYKTAKMLFKLATLAVTAMTKDPVVVVVLIADF-GLE-VFDTGFFSFYQEKLOP 1100

Qy 142 SYAFTYSSKRAIALEHVKLSNRKSTDDYQITLDVV-FDSLEDVGATGFPRTY-ESVEQ 199

Db 1101 YMKTIPIKSELVTDATAAAQIPK-GVYSFVSS 1133

Qy 200 FMSAVGCTNDE-IARLPTSAA-ISKLSYDYNLIPG 231

RESULT 8

ID Q88978 PRELIMINARY; PRT; 2497 AA.

AC Q88978;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE NONSTRUCTURAL POLYPROTEIN.

OS VENEZUELAN EQUINE ENCEPHALITIS VIRUS.

OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; TOGAVIRIDAE;

OC ALPHAVIRUS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-680201 (SUBTYPE IE);

RX MEDLINE; 96204604.

RA OBERSTE M.S., PARKER M.D., SMITH J.F.;

RT "Complete sequence of Venezuelan equine encephalitis virus subtype IE reveals conserved and hypervariable domains within the C terminus of nsP3.";

RL VIROLOGY 219:314-320(1996).

DR EMBL; U34999; G1144528; -

DR PFAM; PF00978; RNA_dep_RNapol2; 2.

KW POLYPROTEIN.

SQ SEQUENCE 2497 AA; 278720 MW; C746FA71 CRC32;

Query Match 4.4%; Score 113; DB 14; Length 2497;

Best Local Similarity 25.7%; Pred. No. 4.22e-02;

Matches 19; Conservative 23; Mismatches 27; Indels 5; Gaps 5;

Db 1125 LRNYDPRINLVPNRRALPHALVTQHADHPSPDFAFVSKL-KGRTVLV-GEKMSISGKT 1182

Qy 62 LRVFDESTGAL-YDSK-SYAFSTSN-DTTSAAAFVSMNSLTNNRIVAILTSGKNFPEV 118

Db 1183 VDWLSETPDSTFRA 1196

Qy 119 VSWLRTAGTSAFPS 132

RESULT 9

ID O43424 PRELIMINARY; PRT; 1007 AA.

AC O43424;

RESULT 11
ID C08906 PRELIMINARY: PRT: 1300 AA.
AC C08906;
DT 01-JUL-1997 (TREMBREL. 04, CREATED)
DT 01-JUL-1997 (TREMBREL. 04, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE REVERSE TRANSCRIPTASE (FRAGMENT).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHA; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN-ORLEANS REELER; TISSUE-LIVER;
RX MEDLINE; 96414308.
RA TAKAHARA T., OHSUMI T., KUROMITSU J., SHIBATA K., SASAKI N.,
RA OKAZAKI Y., SHIBATA H., SATO S., YOSHIMI A., KUSAKABE M.,
RA MURAMATSU M., UEKI M., OKUDA K., HAYASHIZAKI Y.;
RT "Dysfunction of the Orleans reeler gene arising from exon skipping
RT due to transposition of a full-length copy of an active L1 sequence
RT into the skipped exon.";
RL HUM. MOL. GENET. 5:989-993(1996).
DR ENBL; D84391; D1021252; -.
DR PFAM; PF00078; rvt; 1.
DR PFAM; PF01260; AP_endonuclease1; 1.
KW RNA-DIRECTED DNA POLYMERASE.
FT NON_TER 1
SQ SEQUENCE 1300 AA; 152068 MW; 3656A0F1 CRC32;

Query Match 4.2%; Score 106; DB 11; Length 1300;
Best Local Similarity 27.1%; Pred. No. 3.55e-01;
Matches 23; Conservative 22; Mismatches 38; Indels 2; Gaps 2

```

QY 96 NSLUNNRNVAILTSGKVNFPPEVWSWLRCTAGTSAPPSDSILSRFDVSYAAAFYTSSKRAIA 155
      : : : : :
Db 150 FTKETLVKLUKAHIAPHTIIVGENT 174
      : : : : :
QY 156 LEHVLSNRKSTDDYTQITLVDVFD 180
      : : : : :

RESULT 12
ID O39818 PRELIMINARY; PRT; 2247 AA.
AC O39818:
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POLYPROTEIN.
OS EQUINE RHINOVIRUS 1.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE: 96165505.
RT Li F., BROWNING G.F., STUDDERT M.J., CRABB B.S.;
RA "Equine rhinovirus 1 is more closely related to foot-and-mouth
RL disease virus than to other picornaviruses."
RT PROC. NATL. ACAD. SCI. U.S.A. 93:990-995(1996).
DR EMBL: L43052; G2231134; --
DR PFAM; PF00073; rhv; 1.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
KW POLYPROTEIN.
SQ SEQUENCE 2247 AA; 249301 MW; 2DD1PF630 CRC32;

Query Match 4.2%; Score 107; DB 14; Length 2247;
Best Local Similarity 23.4%; Pred. No. 2.62e-01;
Matches 22; Conservative 33; Mismatches 32; Indels 7; Gaps 7

Db 1041 TYSPTPKVAKMLFKLATLAVAAAMRTKPVVVVMLIADF-GLE-VFDTGFFESVFQEKLP 1098
      : : : : :
Ov 142 SYAAEYTSKRAIALFHVKISNPSTQVOTIIVNV-FSLFVCATCTGCTGTV-FQVDA 109
      : : : : :

```



```
RP SEQUENCE FROM N.A.
RC STRAIN-TRINIDAD DONKEY;
RA KINNEY R.;
RL SUBMITTED (AUG-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; L01442; G3201595; -.
DR PFAM; PF00978; RNA_dep_RNapol2; 2.
KW POLYPROTEIN.
FT CHAIN 1 535 NONSTRUCTURAL PROTEIN NSP1.
FT CHAIN 536 1329 NONSTRUCTURAL PROTEIN NSP2.
FT CHAIN 1330 1886 NONSTRUCTURAL PROTEIN NSP3.
FT CHAIN 1887 2493 NONSTRUCTURAL PROTEIN NSP4.
SQ SEQUENCE 2493 AA; 277915 MW; A6B7B200 CRC32;

Query Match 4.2%; Score 106; DB 14; Length 2493;
Best Local Similarity 24.2%; Pred. No. 3.55e-01;
Matches 23; Conservative 26; Mismatches 39; Indels 7; Gaps 7;

Db 1125 LRNYDPRINLPVNNRRLPHALVLHNEHPOSDFSSFSVKL-KGRTVLVY-GEKLSVPGKM 1182
QY 62 LRVFDPSTGAL-VDSK-SYAFSTN-NDTTSAAFYSEFMNSLTNNRIVAILTSGKVNFPPEV 118
Db 1183 VDWLSDRPEATFRARLDLIGPDVDPKYDIIFVNVNR 1217
QY 119 VSWLRTAGTSAPFSDSILS-RFDVS-YAIFYTSSK 151

Search completed: Wed Sep 1 16:02:23 1999
Job time : 64 secs.
```

 W P S R E L E A

 (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed Sep 1 16:06:41 1999; MasPar time 4.33 Seconds
 Tabular output not generated.
 117.961 Million cell updates/sec

Title: >PCT-US99-13024-2
 Description: (1-24) from PCTUS9913024.pap (2 of 12)
 Perfect Score: 176
 Sequence: 1 MEKFMFEFGQGVQTPFLSESNSV 24
 Scoring table: PAM 150
 Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-genseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33
 34:part34 35:part35 36:part36 37:part37 38:part38
 39:part39

Statistics: Mean 21.293; Variance 79.022; scale 0.269

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	74	42.0	65	19	R97371	Phage T4 ORF5 gene pr
2	65	36.9	591	37	W72912	Mycobacterium tuber
3	65	36.9	621	29	W55645	H. pylori ORF 06p106
4	65	36.9	637	30	W55739	H. pylori ORF 06p106
5	65	36.9	645	37	W73030	Helicobacter pylori 7
6	65	36.9	652	37	W73029	Helicobacter pylori 7
7	64	36.4	358	35	W44826	Pseudomonas fluoresce
8	64	36.4	358	35	W44826	Pseudomonas fluoresce
9	63	35.8	364	33	W62047	Streptococcus pneumon
10	63	35.8	389	2	P70668	D-alanine racemase.
11	63	35.8	409	9	R48060	Sequence of protease
12	63	35.8	461	30	W39928	Human CDP-diacylglyce
13	63	35.8	740	33	W62045	Streptococcus pneumon
14	62	35.2	146	29	W55354	H. pylori ORF 01cp117
15	62	35.2	269	29	W55318	H. pylori ORF 07ep119
16	62	35.2	479	29	W55355	H. pylori ORF 02ge101

17	62	35.2	479	29	W55556	H. pylori ORF 02ge101	1.04e+02
18	62	35.2	486	29	W55502	H. pylori ORF 02gp207	1.04e+02
19	61	34.7	43	8	R40095	Hib OMP P1-P2 hybrid	1.26e+02
20	61	34.7	357	29	W55533	H. pylori ORF hp6e203	1.26e+02
21	61	34.7	382	29	W55369	H. pylori ORF 04gp112	1.26e+02
22	61	34.7	451	12	R63088	Breast tumor kinase,	1.26e+02
23	61	34.7	479	29	W55693	H. pylori ORF 09cp107	1.26e+02
24	61	34.7	479	29	W55570	H. pylori ORF 06cp306	1.26e+02
25	61	34.7	479	29	W55569	H. pylori ORF 06cp306	1.26e+02
26	61	34.7	479	29	W55702	H. pylori ORF 09cp107	1.26e+02
27	61	34.7	510	26	W34561	Pyrococcus furiosus g	1.26e+02
28	61	34.7	510	11	R58801	Hyperthermostable be	1.26e+02
29	61	34.7	510	17	R88094	Hyperthermostable bet	1.26e+02
30	61	34.7	510	36	W49865	Pyrococcus VC1-7G1 ql	1.26e+02
31	60	34.1	85	29	W55416	H. pylori ORF hp1p105	1.54e+02
32	60	34.1	130	36	R80717	S. pneumoniae protein	1.54e+02
33	60	34.1	313	37	W75159	Human secreted protei	1.54e+02
34	60	34.1	390	23	W21678	Haemophilus influenza	1.54e+02
35	60	34.1	435	29	W55642	H. pylori ORF 05ep108	1.54e+02
36	60	34.1	451	5	R27642	Human calcium channel	1.54e+02
37	60	34.1	2262	31	W56737	Calcium ion channel a	1.54e+02
38	59	33.5	38	35	W78736	Leader peptide sequen	1.87e+02
39	59	33.5	241	34	W38602	Streptococcus pneumon	1.87e+02
40	59	33.5	380	35	W69360	Glycerol dehydrogenas	1.87e+02
41	59	33.5	438	28	W37355	Wzy (Rfc) protein inv	1.87e+02
42	59	33.5	946	26	W27085	Human cytomegalovirus	1.87e+02
43	59	33.5	1202	8	R44507	BamMV coat protein.	1.87e+02
44	59	33.5	2938	11	R59223	GAP protein Irai.	1.87e+02
45	59	33.5	2942	25	W22050	Saccharomyces cerevis	1.87e+02

ALIGNMENTS

RESULT 1
 ID R97371 standard; Protein; 65 AA.

AC R97371;
 DT 07-JAN-1997 (first entry)
 DE Phage T4 ORF5 gene product, vgp34.
 KW Phage T4; tail fibre protein; nanotechnology; nano-structure;
 KW filter; molecular sieve.
 OS Bacteriophage T4.
 PN W09611947-A1.
 PD 25-APR-1996.
 PF 13-OCT-1995; U13023.
 PR 13-OCT-1994; US-322760.
 PA (GOLD/) GOLDBERG E B.
 PI Goldbergs EB;
 DR WPI; 96-221942/22.
 DR N-PSDB; T29053.
 PT New proteins derived from T4 phage tail fibre proteins - that can
 PT self assemble into nano-structure(s), useful as filters etc, also
 PT corresponding DNA
 PS Claim 7; Fig 7; 83pp; English.
 CC A protein (R97371) of unspecified function is the product of open
 CC reading frame x of the tail fibre protein gene region (see also
 CC T29053) of phage T4. This gene region also includes open reading
 CC frames for tail fibre proteins (see also R97370 and R97372-74).
 CC Tail fibre proteins (native or modified) can be produced in large
 CC quantities in microbial cells and used as building blocks of strong,
 CC stable nanostructures.
 SQ Sequence 65 AA;

Query Match 42.0%; Score 74; DB 19; Length 65;
 Best Local Similarity 58.3%; Pred. No. 8.97e+00;
 Matches 14; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

DB 1 mekfmaefgqgvqtpflsesns 24
 QY 1 MEKFMFEFGQGVQTPFLSESNSV 24

RESULT 2
 ID W72912 standard; Protein; 591 AA.

AC W72912;
 DT 21-JAN-1999 (first entry)
 DE Mycobacterium tuberculosis antigen RDI-ORF2.
 KW Mycobacterium tuberculosis; antigen; vaccine; immunological;
 OS Immunogen; infection.
 PN WO9844119-A1.
 PD 08-OCT-1998.
 PF 01-APR-1998; DK0132.
 PR 05-JAN-1998; US-070488.
 PR 02-APR-1997; DK-000376.
 PR 18-APR-1997; US-044624.
 PR 10-NOV-1997; DK-001277.
 PA (STAT-) STATENS SERUM INST.
 PI Andersen P, Florio W, Nielsen R, Oettinger T, Rasmussen PB,
 PI Rosenkrands J, Weldingh K;
 DR WPI: 98-542705/45.
 DR N-PSDB: V63936.
 PT New isolated mycobacteria polypeptides and nucleic acids - used for
 PT developing products for the diagnosis of or vaccination against
 PT mycobacterial infections, particularly tuberculosis
 PS Claim 1: Page 188-190; 163pp; English.
 CC The present sequence represents a Mycobacterium tuberculosis protein.
 CC Products from the present invention, which describes protein fragments
 CC and nucleic acid fragments derived from M.tuberculosis, can be used in
 CC the detection of and prevention of mycobacterial infections. In
 CC particular, the proteins and nucleic acids can be used for the diagnosis
 CC of or vaccination against tuberculosis caused by M. tuberculosis,
 CC M. africanum or M. bovis.
 CC Sequence 591 AA;
 SQ

Query Match 36.9%; Score 65; DB 37; Length 591;
 Best Local Similarity 55.0%; Pred. No. 5.69e+01;
 Matches 11; Conservative 2; Mismatches 5; Indels 2; Gaps 2;

Db 522 mdkfvgaafgsg-aptmfsls 540

QY 1 MEKFM-AEFGQGVQTPFLS 19

RESULT 3

ID W55645 standard; Protein: 621 AA.
 AC W55645;
 DT 24-JUN-1998 (first entry)
 DE H. pylori ORF 06p10615_961562_f2_41 cell envelope OMP.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacteria; life cycle; activator;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
 OS Helicobacter pylori.
 PN WO9737044-A1.
 PD 09-OCT-1997.
 PF 27-MAR-1997; U05223.
 PR 06-DEC-1996; US-761318.
 PR 29-MAR-1996; US-625811.
 PR 02-APR-1996; US-758731.
 PR 25-OCT-1996; US-736905.
 PR 28-OCT-1996; US-738859.
 PA (ASTR) ASTRA AB.
 PI Alm RA, Smith D;
 DR WPI: 97-503122/46.
 DR N-PSDB: V25054.
 PT Helicobacter pylori nucleic acid sequences and encoded
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
 PT infection and for diagnosis of H. pylori infection
 PS Claims 14, 80; Pages 850-851; 1145pp; English.
 CC This is a H. pylori cell envelope outer membrane protein having a C-
 CC terminal tyrosine cluster motif. The protein may be used in a vaccine to
 CC prevent or treat H. pylori infection or to identify H. pylori polypeptide
 CC binding compounds, useful as potential H. pylori life cycle activators or
 CC inhibitors. The DNA and probes derived from it may be used for the
 CC identification of H. pylori in a sample and the diagnosis of H. pylori
 CC infection. Nucleic acid sequences complementary to the DNA act as
 CC antisense sequences and can be used to prevent the translation of

H. pylori mRNA. Antibodies against the protein can be used in
 CC immunoassays to evaluate the abundance and distribution of H. pylori-
 CC specific antigens. The genomic sequence of H. pylori (ATCC 55679) was
 CC determined from overlapping contigs generated by mechanically shearing
 CC the bacterial DNA. The sequences were analysed for ORF of at least 180
 CC nucleotides, and the predicted coding regions defined by computer
 CC evaluation. To identify likely H. pylori antigens for vaccine
 CC development, the amino acid sequences predicted from various ORF were
 CC analysed for significant homology to other known or exported membrane
 CC proteins. Having identified and determined the sequences of interest,
 CC particular regions can be isolated from H. pylori by PCR amplification
 CC for recombinant polypeptide production, e.g. in E. coli hosts.
 SQ Sequence 621 AA;

Query Match 36.9%; Score 65; DB 29; Length 621;
 Best Local Similarity 22.2%; Pred. No. 5.69e+01;
 Matches 4; Conservative 10; Mismatches 4; Indels 0; Gaps 0;

Db 475 dynhgykssfnssd1 492

QY 7 EFGQGVQTPFLSESNSV 24

RESULT 4

ID W55739 standard; Protein: 637 AA.
 AC W55739;
 DT 13-JUL-1998 (first entry)
 DE H. pylori ORF 06p10615_961562_f1_15 cell envelope OMP.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacteria; life cycle; activator;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
 KW cell envelope outer membrane protein; OMP; tyrosine cluster motif.
 OS Helicobacter pylori.
 PN WO9737044-A1.
 PD 09-OCT-1997.
 PF 27-MAR-1997; U05223.
 PR 06-DEC-1996; US-761318.
 PR 29-MAR-1996; US-625811.
 PR 02-APR-1996; US-758731.
 PR 25-OCT-1996; US-736905.
 PR 28-OCT-1996; US-738859.
 PA (ASTR) ASTRA AB.
 PI Alm RA, Smith D;
 DR WPI: 97-503122/46.
 DR N-PSDB: V25148.
 PT Helicobacter pylori nucleic acid sequences and encoded
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
 PT infection and for diagnosis of H. pylori infection
 PS Claims 14, 80; Pages 1108,1109; 1145pp; English.
 CC This sequence represents a Helicobacter pylori cell envelope outer
 CC membrane protein having a C-terminal tyrosine cluster motif.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors. The
 CC DNA and probes derived from it may be used for the identification of
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
 CC acid sequences complementary to the DNA act as antisense sequences and
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies
 CC against the protein can be used in immunoassays to evaluate the abundance
 CC and distribution of H. pylori-specific antigens. The genomic sequence of
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were analysed
 CC for ORF of at least 180 nucleotides, and the predicted coding regions
 CC defined by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences of
 CC interest, particular regions can be isolated from H. pylori by PCR
 CC amplification for recombinant polypeptide production, e.g. in E. coli
 CC hosts.
 SQ Sequence 637 AA;

Query Match 36.9%; Score 65; DB 30; Length 637;

PN	WO9843479-A1.	08-OCT-1998.	U06421.	31-MAR-1998; US-834666.	01-APR-1997; US-831310.	01-APR-1997; US-831310.	(HUMA-) HUMAN GENOME SCI INC.	PI (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS ET VACCINS.	PI Al-Garawi A, Kleanthous H, Lissolo L, Miller C, Tomb J; WPI: 98-568251/48.	DR N-PSDB: V07918.	PT New isolated Helicobacter polynucleotides - used to develop products for the diagnosis, prevention and treatment of Helicobacter infections and gastroduodenal diseases	PT Claim 9; Page 117-120; 184pp; English.	CC This is the amino acid sequence of a 76 kDa Helicobacter pylori polypeptide designated GHPO 190. It was deduced from an isolated genomic DNA sequence (see V07918). The invention provides a family of 76 kDa Helicobacter polypeptides (see W73022-32), as well as a 32 kDa polypeptide (see W73034) and a 50 kDa polypeptide (see W73035), and also polynucleotides (see V72001, V07912-21 and V07963-64) encoding them, expression cassettes, and methods for producing the unprocessed or mature polypeptides in host cells. The polypeptides can be used in vaccination methods to prevent or treat Hb infection in a mammal. Methods and products of the invention allow treatment and prevention of gastroduodenal diseases associated with Hb infections, including acute, chronic, and atrophic gastritis, and peptic ulcer diseases, e.g. gastric and duodenal ulcers. Detection and diagnostic methods are also provided. GHPO 190 was demonstrated to be a protective antigen.	CC Sequence 652 AA;	SO
Query Match	36.9%;	Score 65;	DB 37;	Length 652;	Best Local Similarity 22.2%;	Pred. No. 5.69e+01;	Mismatches 4;	Conservative 10;	Mismatches 4;	Indels 0;	Gaps 0;				
Db	506 dnhgykssffnssdi 523	7 EFGQGYVOTPFLESNSV 24	506 dnhgykssffnssdi 523	7 EFGQGYVOTPFLESNSV 24	506 dnhgykssffnssdi 523	7 EFGQGYVOTPFLESNSV 24	506 dnhgykssffnssdi 523	7 EFGQGYVOTPFLESNSV 24	506 dnhgykssffnssdi 523	7 EFGQGYVOTPFLESNSV 24	506 dnhgykssffnssdi 523	7 EFGQGYVOTPFLESNSV 24			
QY	7 EFGQGYVOTPFLESNSV 24	7 EFGQGYVOTPFLESNSV 24	7 EFGQGYVOTPFLESNSV 24	7 EFGQGYVOTPFLESNSV 24	7 EFGQGYVOTPFLESNSV 24	7 EFGQGYVOTPFLESNSV 24	7 EFGQGYVOTPFLESNSV 24	7 EFGQGYVOTPFLESNSV 24	7 EFGQGYVOTPFLESNSV 24	7 EFGQGYVOTPFLESNSV 24	7 EFGQGYVOTPFLESNSV 24	7 EFGQGYVOTPFLESNSV 24			
RESULT	7	7	7	7	7	7	7	7	7	7	7	7			
ID	W44826;	standard; Protein; 358 AA.	W44826;	standard; Protein; 358 AA.	W44826;	standard; Protein; 358 AA.	W44826;	standard; Protein; 358 AA.	W44826;	standard; Protein; 358 AA.	W44826;	standard; Protein; 358 AA.			
DT	23-NOV-1998	(first entry)	23-NOV-1998	(first entry)	23-NOV-1998	(first entry)	23-NOV-1998	(first entry)	23-NOV-1998	(first entry)	23-NOV-1998	(first entry)			
DE	Pseudomonas fluorescens strain P.J.874 HPPD protein.	Transgenic plant; resistance; herbicide; isoxaflutole; sulcotriene; primer; hydroxyphenylpyruvate dioxygenase; tobacco; cotton; rape; sugar beet; cereal; toxinyl; nitrilase; dihalohydroxybenzotrile; bromoxynil; nematode; insecticide; glyphosate oxidoreductase; Bacillus thuringiensis; fungus.	Pseudomonas fluorescens strain P.J.874 HPPD protein.	Transgenic plant; resistance; herbicide; isoxaflutole; sulcotriene; primer; hydroxyphenylpyruvate dioxygenase; tobacco; cotton; rape; sugar beet; cereal; toxinyl; nitrilase; dihalohydroxybenzotrile; bromoxynil; nematode; insecticide; glyphosate oxidoreductase; Bacillus thuringiensis; fungus.	Pseudomonas fluorescens strain P.J.874 HPPD protein.	Transgenic plant; resistance; herbicide; isoxaflutole; sulcotriene; primer; hydroxyphenylpyruvate dioxygenase; tobacco; cotton; rape; sugar beet; cereal; toxinyl; nitrilase; dihalohydroxybenzotrile; bromoxynil; nematode; insecticide; glyphosate oxidoreductase; Bacillus thuringiensis; fungus.	Pseudomonas fluorescens strain P.J.874 HPPD protein.	Transgenic plant; resistance; herbicide; isoxaflutole; sulcotriene; primer; hydroxyphenylpyruvate dioxygenase; tobacco; cotton; rape; sugar beet; cereal; toxinyl; nitrilase; dihalohydroxybenzotrile; bromoxynil; nematode; insecticide; glyphosate oxidoreductase; Bacillus thuringiensis; fungus.	Pseudomonas fluorescens strain P.J.874 HPPD protein.	Transgenic plant; resistance; herbicide; isoxaflutole; sulcotriene; primer; hydroxyphenylpyruvate dioxygenase; tobacco; cotton; rape; sugar beet; cereal; toxinyl; nitrilase; dihalohydroxybenzotrile; bromoxynil; nematode; insecticide; glyphosate oxidoreductase; Bacillus thuringiensis; fungus.	Pseudomonas fluorescens strain P.J.874 HPPD protein.	Transgenic plant; resistance; herbicide; isoxaflutole; sulcotriene; primer; hydroxyphenylpyruvate dioxygenase; tobacco; cotton; rape; sugar beet; cereal; toxinyl; nitrilase; dihalohydroxybenzotrile; bromoxynil; nematode; insecticide; glyphosate oxidoreductase; Bacillus thuringiensis; fungus.			
OS	Pseudomonas fluorescens.		Pseudomonas fluorescens.		Pseudomonas fluorescens.		Pseudomonas fluorescens.		Pseudomonas fluorescens.		Pseudomonas fluorescens.				
PN	WO9802562-A2.		WO9802562-A2.		WO9802562-A2.		WO9802562-A2.		WO9802562-A2.		WO9802562-A2.				
PD	22-JAN-1998.		22-JAN-1998.		22-JAN-1998.		22-JAN-1998.		22-JAN-1998.		22-JAN-1998.				
PF	10-JUL-1997; F01256.		10-JUL-1997; F01256.		10-JUL-1997; F01256.		10-JUL-1997; F01256.		10-JUL-1997; F01256.		10-JUL-1997; F01256.				
PA	16-JUL-1996; FR-009137.		16-JUL-1996; FR-009137.		16-JUL-1996; FR-009137.		16-JUL-1996; FR-009137.		16-JUL-1996; FR-009137.		16-JUL-1996; FR-009137.				
PI	Derose R, Pallett K, Pellissier B, Sailland A;		Derose R, Pallett K, Pellissier B, Sailland A;		Derose R, Pallett K, Pellissier B, Sailland A;		Derose R, Pallett K, Pellissier B, Sailland A;		Derose R, Pallett K, Pellissier B, Sailland A;		Derose R, Pallett K, Pellissier B, Sailland A;				
DR	WPI: 98-110608/10.		WPI: 98-110608/10.		WPI: 98-110608/10.		WPI: 98-110608/10.		WPI: 98-110608/10.		WPI: 98-110608/10.				
DR	N-PSDB: V19280.		N-PSDB: V19280.		N-PSDB: V19280.		N-PSDB: V19280.		N-PSDB: V19280.		N-PSDB: V19280.				
PT	Chimaeric gene imparting resistance to two or more herbicides - such as hydroxyphenylpyruvate dioxygenase inhibitors, also related vectors and transformed plants		Chimaeric gene imparting resistance to two or more herbicides - such as hydroxyphenylpyruvate dioxygenase inhibitors, also related vectors and transformed plants		Chimaeric gene imparting resistance to two or more herbicides - such as hydroxyphenylpyruvate dioxygenase inhibitors, also related vectors and transformed plants		Chimaeric gene imparting resistance to two or more herbicides - such as hydroxyphenylpyruvate dioxygenase inhibitors, also related vectors and transformed plants		Chimaeric gene imparting resistance to two or more herbicides - such as hydroxyphenylpyruvate dioxygenase inhibitors, also related vectors and transformed plants		Chimaeric gene imparting resistance to two or more herbicides - such as hydroxyphenylpyruvate dioxygenase inhibitors, also related vectors and transformed plants				
PS	Disclosure; Fig 1: 49pp; French.		Disclosure; Fig 1: 49pp; French.		Disclosure; Fig 1: 49pp; French.		Disclosure; Fig 1: 49pp; French.		Disclosure; Fig 1: 49pp; French.		Disclosure; Fig 1: 49pp; French.				
CC	The invention relates to the generation of transgenic plants resistant to preferably at least 2 herbicides by introducing a chimaeric gene comprising at least 2 gene subunits where one sequence encodes hydroxyphenylpyruvate dioxygenase (HPPD), an enzyme that promotes tolerance to herbicides such as isoxaflutole or sulcotriene. This sequence represents the HPPD protein from Pseudomonas fluorescens strain P.J.874. The sequence was used to derive the degenerate coding sequence (V19280). This in turn was used to design the degenerate primers (V19270-V19274) for the isolation of the HPPD gene from P. fluorescens strain A32 (V19269). The plants, e.g. tobacco, cotton, rape, soya,		The invention relates												


```

RESULT 11
ID R48060 standard; Protein; 409 AA.
AC R48060;
DE 20-JUL-1994 (first entry)
DT Sequence of protease A encoded by gene PRA1.
KW Protease; Yeast; proteolysis.
OS Kluyveromyces lactis.
PN W09400578-A.
PD 06-JAN-1994.
PF 23-JUN-1993; F00623.
PR 25-JUN-1992; FR-007785.
PA (RHON ) RHONE POULENC RORER SA.
PI Fleer R, Fournier A, Yeh P;
DR WPI; 94-026215/03.
DR N-PSDB: 055348.
PT New Kluyveromyces yeast with modified protease gene - esp. used
PT for high yield prodn. of recombinant protein, also DNA encoding
PT yeast protease and derived peptide(s).
PS Disclosure: Page 31-33; 49pp; English.
CC The protease gene is to be modified in order to render it (partially)
CC incapable of producing the natural protein; or result in a non-
CC functional protease or in a protease with modified proteolytic
CC activity. The modifications can be introduced in vitro or in situ by
CC standard genetic engineering techniques or by exposure to mutagenic
CC agents.
SQ Sequence 409 AA;

Query Match 35.8%; Score 63; DB 9; Length 409;
Best Local Similarity 30.4%; Pred. NO. 8.50e+01;
Matches 7; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

Db 54 ekylnsfgraypqesfskdnvdv 76
|:::|:::|:::|:::|:::|
Qy 2 EKFMAEFGGYVQTPFLSENSV 24

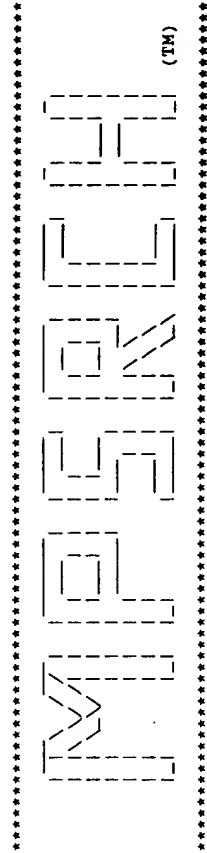
RESULT 12
ID t13928 standard; Protein; 461 AA.
AC W39928;
DE 08-JUL-1998 (first entry)
DT Human CDP-diacylglycerol synthase.
DE CDP-diacylglycerol synthase; hCDS; human;
KW CTP:phosphatidate cytidyltransferase; antiinflammatory;
KW haematopoiesis; reoxygenation injury; therapy.
OS Homo sapiens.
PF Key
FH Location/Qualifiers
FT Misc_difference 412
FT /note= "encoded by GGN"
FT W09800550-A1.
PN 08-JAN-1998.
PD 25-JUN-1997; U11294.
PF 28-JUN-1996; US-672814.
PA (CELL-) CELL THERAPEUTICS INC.
PI Leung DW, Weeks R;
PI WPI; 98-086975/08.
DR N-PSDB: V09944.
PT Nucleic acid encoding human CDP-diacyl-glycerol synthase and
PT related proteins - used to identify compounds that are
PT antiinflammatory, increase haematopoiesis or prevent re-oxygenation
PT injury
PS Claim 1; Page 25-26; 56pp; English.
CC This polypeptide comprises human CDP-diacylglycerol synthase
CC (hCDS), or CTP:phosphatidate cytidyltransferase, an enzyme that
CC converts phosphatidic acid to CDP-diacyl-glycerol, a precursor of
CC phosphatidylinositol, phosphatidylglycerol and cardiolipin. Its
CC amino acid sequence was deduced from an isolated cDNA clone (see
CC W09944). It shows homology to Drosophila CDS (see W39929), yeast
CC CDS and E. coli CDS. hCDS polypeptides can be obtained by chemical
CC synthesis or by recombinant expression. A claimed method for
CC screening drug candidate compounds having activity as antiinflammatory
CC agents, for increasing haematopoiesis, and preventing reoxygenation
CC injury following cytoreductive therapy involves contacting a CDS

```


KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacteria; life cycle; activator;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
 OS Helicobacter pylori.
 PN W09737044-A1.
 PD 09-OCT-1997; U05223.
 PF 27-MAR-1997; US-761318.
 PR 06-DEC-1996; US-761318.
 PR 29-MAR-1996; US-625811.
 PR 02-APR-1996; US-758731.
 PR 25-OCT-1996; US-736905.
 PR 28-OCT-1996; US-738859.
 PA (ASTR) ASTRA AB.
 PI Alm RA, Smith D;
 DR WPI; 97-503122/46.
 DR N-PSDB; V24764.
 PT Helicobacter pylori nucleic acid sequences and encoded
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
 PT infection and for diagnosis of H. pylori infection
 PS Claim 14; Page 576-577; 1145pp; English.
 CC This sequence is a H. pylori protein of unspecified function.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors. The
 CC DNA and probes derived from it may be used for the identification of
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
 CC acid sequences complementary to the DNA act as antisense sequences and
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies
 CC against the protein can be used in immunoassays to evaluate the abundance
 CC and distribution of H. pylori-specific antigens. The genomic sequence of
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were analysed
 CC for ORF of at least 180 nucleotides, and the predicted coding regions
 CC defined by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences of
 CC interest, particular regions can be isolated from H. pylori by PCR
 CC amplification for recombinant polypeptide production, e.g. in E. coli
 CC hosts.
 SQ Sequence 146 AA;
 Query Match 35.2%; Score 62; DB 29; Length 146;
 Best Local Similarity 45.0%; Pred. No. 1.04e+02;
 Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
 Db 35 etfvtavqggniyadflsks 54
 | | | | | | | | | |
 QY 2 EKFWAEFGQGVQTPFLSES 21
 | | | | | | | | | |
 RESULT 15
 ID W55518 standard; Protein; 269 AA.
 AC W55518;
 DT 02-JUL-1998 (first entry)
 DE H. pylori ORF 07ep11916_5913592_f3_18 cell envelope OMP.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacteria; life cycle; activator;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
 OS Helicobacter pylori.
 PN W09737044-A1.
 PD 09-OCT-1997; U05223.
 PF 27-MAR-1997; US-761318.
 PR 06-DEC-1996; US-761318.
 PR 29-MAR-1996; US-625811.
 PR 02-APR-1996; US-758731.
 PR 25-OCT-1996; US-736905.
 PR 28-OCT-1996; US-738859.
 PA (ASTR) ASTRA AB.
 PI Alm RA, Smith D;
 DR WPI; 97-503122/46.
 DR N-PSDB; V24927.
 PT Helicobacter pylori nucleic acid sequences and encoded

PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
 PT infection and for diagnosis of H. pylori infection
 PS Claims 14,80; Page 721-722; 1145pp; English.
 CC This sequence is a H. pylori cell envelope outer membrane protein which
 CC is stated to have a terminal Phe residue.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors. The
 CC DNA and probes derived from it may be used for the identification of
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
 CC acid sequences complementary to the DNA act as antisense sequences and
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies
 CC against the protein can be used in immunoassays to evaluate the abundance
 CC and distribution of H. pylori-specific antigens. The genomic sequence of
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were analysed
 CC for ORF of at least 180 nucleotides, and the predicted coding regions
 CC defined by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences of
 CC interest, particular regions can be isolated from H. pylori by PCR
 CC amplification for recombinant polypeptide production, e.g. in E. coli
 CC hosts.
 SQ Sequence 269 AA;
 Query Match 35.2%; Score 62; DB 29; Length 269;
 Best Local Similarity 45.0%; Pred. No. 1.04e+02;
 Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
 Db 184 etfvtavqggniyadflsks 203
 | | | | | | | | | |
 QY 2 EKFWAEFGQGVQTPFLSES 21
 | | | | | | | | | |

Search completed: Wed Sep 1 16:07:08 1999
 Job time : 27 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:06:10 1999; MasPar time 4.18 Seconds
Tabular output not generated. 229.968 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (1-24) from PCTUS9913024.pep (2 of 12)
Perfect Score: 176
Sequence: 1 MEKPMAEFGQGYVOTPLSENSV 24

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 30.660; Variance 51.855; scale 0.591

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	77	43.8	850	2	F71000 hypothetical protein	1.61e-01
2	73	41.5	2301	2	T02323 hypothetical protein	7.07e-01
3	68	38.6	554	2	S41574 DNA ligase (ATP) (EC	4.22e+00
4	68	38.6	809	1	Q0B834 BBLF4 protein - human	4.22e+00
5	68	38.6	1224	2	A25884 DNA-directed RNA poly	4.22e+00
6	67	38.1	236	2	JQ0061 hypothetical 25K prot	5.97e+00
7	66	37.5	56	2	S35168 cytochrome P450 (clon	8.42e+00
8	66	37.5	70	2	E64757 probable membrane pro	8.42e+00
9	66	37.5	112	1	GMBPT4 gene 33 protein - pha	8.42e+00
10	66	37.5	336	2	B71121 hypothetical protein	8.42e+00
11	66	37.5	337	2	S21209 4-hydroxyphenylpyruva	8.42e+00
12	66	37.5	508	2	I64032 hypothetical protein	8.42e+00
13	66	37.5	518	2	H64775 probable membrane pro	8.42e+00
14	66	37.5	853	2	H70470 hypothetical protein	8.42e+00
15	66	37.5	1191	2	S65068 DNA-directed RNA poly	8.42e+00
16	65	36.9	277	2	S32371 maltodextrin - Strept	1.18e+01
17	65	36.9	591	2	E70802 hypothetical protein	1.18e+01
18	65	36.9	608	2	B64610 outer membrane protei	1.18e+01
19	65	36.9	629	2	E64610 outer membrane protei	1.18e+01
20	65	36.9	638	2	B71904 probable outer membra	1.18e+01
21	65	36.9	651	2	E71904 probable outer membra	1.18e+01
22	64	36.4	182	2	I64029 hypothetical protein	1.66e+01
23	64	36.4	273	2	D70428 hypothetical protein	1.66e+01

24	64	36.4	420	1	TVUTG4	phosphoglycerate kina	1.66e-01
25	64	36.4	440	1	TVUTGB	phosphoglycerate kina	1.66e-01
26	64	36.4	440	1	KIUTGG	phosphoglycerate kina	1.66e-01
27	64	36.4	505	1	TVUT2B	phosphoglycerate kina	1.66e-01
28	64	36.4	508	1	TVUT4B	phosphoglycerate kina	1.66e-01
29	64	36.4	509	1	A45593	phosphoglycerate kina	1.66e-01
30	64	36.4	656	2	S61097	hypothetical protein	1.66e-01
31	64	36.4	980	2	T00045	cellodextrin phosphor	1.66e-01
32	64	36.4	1188	2	S30229	DNA-directed RNA poly	1.66e-01
33	63	35.8	175	2	G71085	hypothetical protein	2.31e-01
34	63	35.8	278	2	A70034	maltodextrin transpor	2.31e-01
35	63	35.8	389	1	JS0443	alanine racemase (EC	2.31e-01
36	63	35.8	389	1	S36653	kpsD protein - Escher	2.31e-01
37	63	35.8	417	1	TVCRGC	phosphoglycerate kina	2.31e-01
38	63	35.8	455	2	E71569	probable acyltransfer	2.31e-01
39	63	35.8	455	1	TVCRGG	phosphoglycerate kina	2.31e-01
40	63	35.8	603	1	W1WL8	E1 protein - human pa	2.31e-01
41	63	35.8	645	2	F70825	probable PPE protein	2.31e-01
42	63	35.8	686	2	S48467	hypothetical protein	2.31e-01
43	63	35.8	730	2	S54625	probable membrane pro	2.31e-01
44	63	35.8	1215	2	S60304	hypothetical protein	2.31e-01
45	62	35.2	402	2	S23860	chloramphenicol resis	3.21e-01

ALIGNMENTS

RESULT 1

ENTRY #type complete

TITLE hypothetical protein PH1302 - Pyrococcus horikoshii

ORGANISM #formal_name Pyrococcus horikoshii

DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998

ACCESSIONS F71000

REFERENCE A71000

#authors Kwarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.

#journal DNA Res. (1998) 5:55-76

#title Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.

#cross-references MOID:98344137

#accession F71000

#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule_type DNA

#residues 1-850 #label KAW

#cross-references GB:AP000006; MID:g3236133; PID:d1031349; PID:g3257723

#experimental_source strain OT3

#note this accession replaces an interim accession for a sequence replaced by GenBank

GENETICS

#gene PH1302

SUMMARY

#length 860 #molecular-weight 95454 #checksum 3711

Query Match 43.8%; Score 77; DB 2; Length 860;

Best Local Similarity 29.4%; Pred. No. 1.61e-01;

Matches 5; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

Db 1 MSKFLDFSNAIRVAF 17

QY 1 MEKPMAEFGQGYVOTPL 17

RESULT 2

ENTRY #type complete

TITLE hypothetical protein F13P17.19 - Arabidopsis thaliana

ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear cress

```

DATE      05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change
ACCESSIONS 05-Mar-1999
REFERENCE  T02323
#authors   Z14168
#submisson submitted to the EMBL Data Library, July 1998
#description Arabidopsis thaliana chromosome II BAC F13P17 genomic
#accession T02323
#status    preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues  1-2301 #label ROU
#cross-references EMBL:AC004481; NID:g3337347; PID:g3337366
GENETICS
#map_position 2
#introns      155/3; 297/1; 525/2; 550/1; 609/1; 767/3; 911/1; 1139/2;
              1238/3; 1412/1; 1498/1; 1528/3; 1598/3; 1732/3; 1791/2;
              1822/3; 1841/3; 1884/3; 1907/3; 1945/3; 1999/3; 2033/3;
              2060/2; 2107/2
#note        F13P17.19
SUMMARY      #length 2301 #molecular-weight 253398 #checksum 4265
Query Match  41.5%; Score 73; DB 2; Length 2301;
Best Local Similarity 50.0%; Pred. No. 7.07e-01;
Matches      8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
Db 977 NFLGRFGAGYVSDTFL 992
      ::::|||||
Qy 3  KFMFAEFGGYVOTPFL 18

RESULT      3
ENTRY       S41974 #type complete
TITLE       DNA ligase (ATP) (EC 6.5.1.1) - fowlpox virus
ORGANISM    #formal_name fowlpox virus
DATE        06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
ACCESSIONS  S41974
REFERENCE    S41971
#authors     Skinner, M.A.; Moore, J.B.; Binns, M.M.; Boursnell, M.E.
#submisson   Submitted to the EMBL Data Library, February 1994
#description Deletion of fowlpox virus homologues of vaccinia virus genes
              between the 3b-hydroxysteroid dehydrogenase (A44L) and DNA
              ligase (A50R) genes.
#accession   S41974
#status      preliminary
#molecule_type DNA
#residues    1-564 #label SK1
#cross-references EMBL:229716; NID:g453598; PID:g453602
CLASSIFICATION #superfamily vaccinia virus DNA ligase
KEYWORDS      ligase; phosphoprotein
FEATURE       236
#active_site Lys (covalent AMP-binding) #status predicted
SUMMARY      #length 564 #molecular-weight 64472 #checksum 7075
Query Match  38.6%; Score 68; DB 2; Length 564;
Best Local Similarity 40.0%; Pred. No. 4.22e+00;
Matches      10; Conservative 8; Mismatches 6; Indels 1; Gaps 1;
Db 79 MYKVVINIGDVAYVIGSLKKSIV 103
Qy 1  MEKFMFAEFGQ-GYVOTPFLSESNSV 24

RESULT      4
ENTRY       Q0BE34 #type complete
TITLE       BBLP4 protein - human herpesvirus 4 (strain B95-8)
ORGANISM    #formal_name human herpesvirus 4, Epstein-Barr virus
#date       25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change

```

```

ACCESSIONS 05-Sep-1997
REFERENCE  F43043; A03776; S33026
#authors   Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
#journal   Mol. Biol. Med. (1983) 1:21-45
#title     Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8
              Epstein-Barr virus.
#cross-references MUID:85035713
#accession F43043
#molecule_type DNA
#residues    1-809 #label BAN
#cross-references EMBL:V01555; NID:g59074; PID:g1334885
REFERENCE  A03794
#authors   Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.;
              Farrell, P.J.; Gibson, T.J.; Hatfull, G.; Hudson, G.S.;
              Satchwell, S.C.; Seguin, C.; Tuffnell, P.S.; Barrell, B.G.
#journal   Nature (1984) 310:207-211
#title     DNA sequence and expression of the B95-8 Epstein-Barr virus
              genome.
#cross-references MUID:84270667
#contents   annotation; protein coding region
CLASSIFICATION #superfamily varicella-zoster virus gene 55 protein
SUMMARY      #length 809 #molecular-weight 89853 #checksum 6134
Query Match  38.6%; Score 68; DB 1; Length 809;
Best Local Similarity 42.1%; Pred. No. 4.22e+00;
Matches      8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
Db 46 METFSPFDPFELSEPPFLP 64
      ::::|||||
Qy 1  MEKFMFAEFGGYVOTPFLS 19

RESULT      5
ENTRY       A25884 #type complete
TITLE       DNA-directed RNA polymerase (EC 2.7.7.6) II 140K chain -
              yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES DNA-directed RNA polymerase B chain B150; protein O3533;
                  protein YOR151c
ORGANISM      #formal_name Saccharomyces cerevisiae
DATE          16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change
ACCESSIONS    A25884; S36860; A37894; S67039
REFERENCE      A25884
#authors       Sweetser, D.; Nonet, M.; Young, R.A.
#journal       Proc. Natl. Acad. Sci. U.S.A. (1987) 84:1192-1196
#title         Prokaryotic and eukaryotic RNA polymerases have homologous
              core subunits.
#cross-references MUID:87147239
#accession     A25884
#molecule_type DNA
#residues      1-1224 #label SWE
#cross-references EMBL:M15693
REFERENCE      S36860
#authors       Sweetser, D.; Nonet, M.; Young, R.A.
#submisson     Submitted to the EMBL Data Library, July 1987
#description    Prokaryotic and eukaryotic RNA polymerases have homologous
              core subunits.
#accession     S36860
#molecule_type DNA
#residues      1-1002,'RRRY',1007-1224 #label SW2
#cross-references EMBL:M15693; NID:g172210; PID:g172211
REFERENCE      A37894
#authors       Riva, M.; Carles, C.; Sentenac, A.; Grachev, M.A.; Mustaev,
              A.A.; Zaychikov, E.F.
#journal       J. Biol. Chem. (1990) 265:16498-16503
#title         Mapping the active site of yeast RNA polymerase B (II).
#cross-references MUID:90375520
#accession     A37894
#molecule_type protein
#residues      961-990 #label RIV
REFERENCE      S67032
#authors       Bordonne, R.; Camasses, A.; Madania, A.; Martin, R.P.; Poch,

```

```

#accession S35168
##molecule_type mRNA
##residues 1-56 ##label MEI
##cross-references EMBL:X69775; NID:g395301; PID:g395302
CLASSIFICATION #superfamily unassigned cytochrome P450; cytochrome P450
homology
KEYWORDS chromoprotein; electron transfer; heme; iron; monooxygenase;
oxidoreductase
FEATURE
2
SUMMARY
    #binding_site heme iron (Cys) (axial ligand) #status
    predicted
    length 56 #checksum 2986
Query Match 37.5%; Score 66; DB 2: Length 56;
Best Local Similarity 29.4%; Pred.No. 8.42e+00;
Matches 5; Conservative 9; Mismatches 3; Indels 0; Gaps 0
Db 20 LQREKFDVAPSYHVPF 36
::|:: : : |::|::|
Qy 1 MEKFMAEFGGQYQTFF 17
RESULT 8
ENTRY E64757 #type complete
TITLE probable membrane protein b0309 - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
24-Oct-1998
ACCESSIONS E64757
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao
Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MIMD:97426617
#accession E64757
#status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-70 ##label BLAT
##cross-references GB:AS000137; GB:U00096; NID:g2367108; PID:g1786500;
UNGP:b0309
#experimental_source strain K-12, substrain MG1655
transmembrane protein
KEYWORDS
FEATURE
34-50 #domain transmembrane #status predicted #label TMM
SUMMARY #length 70 #molecular_weight 7854 #checksum 6854
Query Match 37.5%; Score 66; DB 2: Length 70;
Best Local Similarity 53.3%; Pred.No. 8.42e+00;
Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0
Db 47 QVVYQTSYLSQQSSI 61
| ||||::|::|:|:|
Qy 10 QGYVQTPFLSESNV 24
RESULT 9
ENTRY GWBFT4 #type complete
TITLE gene 33 protein - phage T4
ORGANISM #formal_name phage T4
#note host Escherichia coli
DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
05-Sep-1997
ACCESSIONS S05558
REFERENCE S05555
#authors Hahn, S.; Rueger, W.
#journal Nucleic Acids Res. (1989) 17:6729
#title Organization of the bacteriophage T4 genome between map
positions 150.745 and 145.824.
#cross-references MIMD:89386003

```

[illegible]

polymerase II.
#cross-references MUID:96178872
#accession S65068
##status nucleic acid sequence not shown
##molecule_type mRNA
##residues 1-1191 #label WAR
##cross-references EMBL:U28403; NID:g1049067; PID:g1049068

##cross-references EMBL:U28403; NID:g1049067; PID:g1049068

GENEID
#gene
CLASSIFICATION
KEYWORDS
SUMMARY
Query Match 37.5% Score 66 DB 2: Length 1191:
Best Local Similarity 35.0% Pred NO. 8.42e+00:
RBP2
#superfamily DNA-directed RNA polymerase 132k polypeptide
DNA binding; nucleotidyltransferase; nucleus; transcription
#length 1191 #molecular-weight 135063 #checksum 5669

Matches 7; Conservative 8; Mismatches 5; Indels 0; Gaps 0

Db 85 YKINFGQIYLSKPMMTESDG 104
:::||||:|::|::

QY 4 FMAEFGQYVQTFFLESNS 23

W O E L H (TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Sep 1 16:05:00 1999; MasPar time 2.92 Seconds
Tabular output not generated. 232.136 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (1-24) from PCTUS9913024.pep (2 of 12)
Perfect Score: 176

Sequence: 1 MEKFMAEFGQYVQTPFLSESNSV 24

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss:prot37
1:swissprot

Statistics: Mean 31.468; Variance 47.304; scale 0.665

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description	Pred. No.
1	74	42.0	65	1	Y15A_BPT4	HYPOTHETICAL 7.3 KD PR	1.45e+01
2	70	39.8	923	1	REI1_SCHPO	MEIOTIC RECOMBINATION	7.11e+01
3	68	38.6	564	1	DNLI_FOWPM	DNA LIGASE (EC 6.5.1.1)	1.54e+00
4	68	38.6	809	1	HELI_EBV	PROBABLE HELICASE.	1.54e+00
5	68	38.6	1224	1	RPB2_YEAST	DNA-DIRECTED RNA POLYM	1.54e+00
6	67	38.1	226	1	COAT_MCMV	COAT PROTEIN (CAPSID P	2.25e+00
7	66	37.5	112	1	VG33_BPT4	RNA POLYMERASE-ASSOCIA	3.28e+00
8	66	37.5	357	1	HPD_PSESP	4-HYDROXYPHENYLPYRUVAT	3.28e+00
9	66	37.5	508	1	YF00_HAEN	HYPOTHETICAL PROTEIN H	3.28e+00
10	66	37.5	518	1	YLAB_ECOLI	HYPOTHETICAL 58.9 KD P	3.28e+00
11	66	37.5	524	1	CP72_CATRO	CYTOKROME P450 72A1 (3.28e+00
12	65	36.9	277	1	MALD_STRPN	MALTODEXTRIN TRANSPORT	4.76e+00
13	64	36.4	239	1	YQCB_HAEN	HYPOTHETICAL PROTEIN H	6.88e+00
14	64	36.4	420	1	PGKE_TRYBB	PHOSPHOGLYCERATE KINAS	6.88e+00
15	64	36.4	440	1	PGKC_TRYBB	PHOSPHOGLYCERATE KINAS	6.88e+00
16	64	36.4	505	1	PGKA_TRYBB	PHOSPHOGLYCERATE KINAS	6.88e+00
17	64	36.4	508	1	PGKD_TRYBB	PHOSPHOGLYCERATE KINAS	6.88e+00
18	64	36.4	556	1	INET_YEAST	HYPOTHETICAL 74.8 KD P	6.88e+00
19	64	36.4	759	1	TRPG_SCHPO	ANTHRANILATE SYNTHASE	6.88e+00
20	64	36.4	1188	1	RPB2_ARATH	DNA-DIRECTED RNA POLYM	6.88e+00
21	63	35.8	389	1	KSS5_ECOLI	CAPSULE POLYSACCHARIDE	9.89e+00
22	63	35.8	389	1	ALR_BACSU	ALANINE RACEMASE (EC 5	9.89e+00
23	63	35.8	417	1	PGKB_CRIFA	PHOSPHOGLYCERATE KINAS	9.89e+00

24 63 35.8 455 1 PGKC_CRIFA PHOSPHOGLYCERATE KINAS 9.89e+00
25 63 35.8 603 1 VEL_HPVB8 REPLICATION PROTEIN E1 9.89e+00
26 63 35.8 686 1 YIK5_YEAST FUMARATE REDUCTASE 13 1.42e+01
27 62 35.2 119 1 FRDD_PROVU FUMARATE REDUCTASE 13 1.42e+01
28 62 35.2 402 1 OPDE_PSEAE TRANSCRIPTION REGULATO 1.42e+01
29 62 35.2 417 1 PGKB_LEIMA PHOSPHOGLYCERATE KINAS 1.42e+01
30 62 35.2 421 1 PGKB_TRYBB PHOSPHOGLYCERATE KINAS 1.42e+01
31 62 35.2 479 1 PGKC_LEIMA PHOSPHOGLYCERATE KINAS 1.42e+01
32 62 35.2 555 1 HYFG_ECOLI HYDROGENASE-4 COMPONEN 1.42e+01
33 62 35.2 569 1 HYCE_ECOLI FORMATE HYDROGENLYASE 1.42e+01
34 62 35.2 604 1 VEL_HPVB36 REPLICATION PROTEIN E1 1.42e+01
35 62 35.2 652 1 SSPI_SCHPO SERINE/THREONINE-PROTE 1.42e+01
36 62 35.2 858 1 UBPS_HUMAN UBIQUITIN CARBOXYL-TER 1.42e+01
37 62 35.2 858 1 UBPS_MOUSE UBIQUITIN CARBOXYL-TER 1.42e+01
38 61 34.7 96 1 Y112_SSV1 HYPOTHETICAL 11.2 KD P 2.02e+01
39 61 34.7 297 1 KPRS_MYGE RIBOSE-PHOSPHATE PYROP 2.02e+01
40 61 34.7 476 1 VTDB_RABIT VITAMIN D-BINDING PROT 2.02e+01
41 61 34.7 610 1 CHIT_STRPL CHITINASE 63 PRECURSOR 2.02e+01
42 61 34.7 659 1 HEPA_HSV7J DNA HELICASE/PRIMASE C 2.02e+01
43 61 34.7 676 1 VI08_VACCV PUTATIVE RNA HELICASE 2.02e+01
44 61 34.7 676 1 VI08_VARV PUTATIVE RNA HELICASE 2.02e+01
45 61 34.7 762 1 AMO2_BOVIN COPPER AMINE OXIDASE P 2.02e+01

ALIGNMENTS

RESULT 1
ID Y15A_BPT4 STANDARD; PRT; 65 AA.
AC P39509;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DE HYPOTHETICAL 7.3 KD PROTEIN IN GP34-GP35 INTERGENIC REGION.
GN Y15A OR 34.1.
OS BACTERIOPHAGE T4.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE;
OC T4-LIKE PHAGES.
RN [1]
RP SEQUENCE FROM N.A.
RA KUTTER E.M.;
RL SUBMITTED (NOV-1994) TO THE SWISS-PROT DATA BANK.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 65 AA; 7334 MW; C2D7CE2D CRC32;

Query Match 42.0%; Score 74; DB 1; Length 65;

Best Local Similarity 58.3%; Pred. No. 1.45e+01;

Matches 14; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Db 1 MEKFMAEIWTICPNAILSESNSV 24

QY 1 MEKFMAEFGQYVQTPFLSESNSV 24

RESULT 2

ID REI1_SCHPO STANDARD; PRT; 923 AA.
AC Q92380;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE MEIOTIC RECOMBINATION PROTEIN RECI1.
GN RECI1.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE; 97231330.
RA LI Y.F., NUMATA M., WAHLS W.P., SMITH G.R.;
RT *Region-specific meiotic recombination in Schizosaccharomyces pombe:
RT the rec1 gene.;
RL MOL. MICROBIOL. 23:869-878(1997).

RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +
CC RNA(N).
CC -!- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
CC THIS SUBUNIT IS THE SECOND LARGEST COMPONENT OF RNA POLYMERASE II.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE FOUND IN
CC EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA PRECURSOR,
CC POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE III FOR 5S
CC AND TRNA GENES.
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M15693; G172211; -;
DR EMBL: U55020; G1293711; -;
DR EMBL: Z75059; E252039; -;
DR PIR: A25884; A25884.
DR SGD: L0001676; RPB2.
DR PROSITE: PS01166; RNA_POL_BETA; 1.
DR PFAM: PF00562; RNA_POL_B; 1.
KW DNA-DIRECTED RNA POLYMERASE; TRANSCRIPTION; ZINC; ZINC-FINGER;
KW NUCLEAR PROTEIN.
FT ZN_FING 1163 1185 C4-TYPE (POTENTIAL).
FT CONFLICT 1003 1006 AEG1 -> RRRY (IN REF. 1).
SQ SEQUENCE 1224 AA; 138751 MW; 138751 MW; DA2ED93A CRC32;

Query Match 38.68; Score 68; DB 1; Length 1224;
Best Local Similarity 40.94; Pred. No. 1.54e+00;
Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Db 87 KYEISFGKIVYTPMVNESDGV 108
QY 3 KFWAEFGQGVQTPFLSENSV 24

RESULT 6
ID COAT_MCMV STANDARD; PRT; 236 AA.
AC P11642;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DE COAT PROTEIN (CAPSID PROTEIN).
OS MAIZE CHLOROTIC MOTTLE VIRUS (MCMV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; MACHLOMOVIRUS.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE: 89263725.
RA NUTTER R.C., SCHEETS K., PANGANIBAN L.C., LOMMEL S.A.;
RT "The complete nucleotide sequence of the maize chlorotic mottle virus
RT genome".
CC -!- SIMILARITY: COAT PROTEINS FROM CARMOVIRUSES (CARMV, MNSV, TCV),
CC DIANTHOVIRUSES (RCNMV), SOBEMOVIRUSES (MCMV, SBMV), TOMBUSVIRUSES
CC (AMCV, CNV, CYRSV, AND TBSV), AND TNV ARE RELATED IN THEIR S
CC REGION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL: X14736; G60533; -;
DR PIR: JQ0061; JQ0061.
DR PROSITE: PS00555; ICOSAH_VIR_COAT_S; 1.
DR PFAM: PF00729; Viral_coat; 1.
KW COAT PROTEIN. 1 51 R DOMAIN, INTERACTION WITH RNA.
FT DOMAIN 52 217 S DOMAIN, VIRION SHELL.
FT DOMAIN 218 236 P DOMAIN, PROJECTING.
SQ SEQUENCE 236 AA; 25149 MW; 647462A9 CRC32;

Query Match 38.18; Score 67; DB 1; Length 236;
Best Local Similarity 42.18; Pred. No. 2.25e+00;
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Db 111 EFTADTTSGYISMAFLSD 129
QY 2 EKFWAEFGQGVQTPFLSE 20

RESULT 7
ID VG33_BPT4 STANDARD; PRT; 112 AA.
AC P13338;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE RNA POLYMERASE-ASSOCIATED PROTEIN GP33.
GN 33
OS BACTERIOPHAGE T4.
OC VIRUSES; DSNA VIRUSES, NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE;
OC T4-LIKE PHAGES.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-BK536;
RX MEDLINE: 89386003.
RA HAHN S., RUEGER W.;
RT "Organization of the bacteriophage T4 genome between map positions
RT 150.745 and 145.824".
RL NUCLEIC ACIDS RES. 17:6729-6729(1989).
RN [2]
RN CHARACTERIZATION.
RX MEDLINE: 94148776.
RA WINKELMAN J.W., KASSAVETIS G.A., GEIDUSCHEK E.P.;
RT "Molecular genetic analysis of a prokaryotic transcriptional
RT coactivator: functional domains of the bacteriophage T4 gene 33
RT protein".
RL J. BACTERIOL. 176:1164-1171(1994).
CC -!- FUNCTION: BINDS TO E.COLI RNA POLYMERASE AND IS REQUIRED FOR
CC TRANSCRIPTION OF LATE GENES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X15818; G15213; -;
DR PIR: S05558; GWP74.
SQ SEQUENCE 112 AA; 12830 MW; 5B4A3D25 CRC32;

Query Match 37.58; Score 66; DB 1; Length 112;
Best Local Similarity 46.28; Pred. No. 3.28e+00;
Matches 12; Conservative 4; Mismatches 7; Indels 3; Gaps 2;

Db 44 IERNVEEFGMSYLEATTAFLSE-NSI 68
QY 1 MEKFWAEFGQGVQ--TPFLSENSV 24

RESULT 8
ID HPPD_PSESP STANDARD; PRT; 357 AA.

[illegible]

DR PDB; 13PK; 24-DEC-97.
DR PDB; 16PK; 25-NOV-98.
DR PROSITE; PS00111; GLYCERATE_KINASE; 1.
DR PFAM; PF00162; PK; 1.
KW TRANSFERASE: KINASE; GLYCOLYSIS; MULTIGENE FAMILY; GLYCOSOME;
3D-STRUCTURE.
FT VARIANT 75 75 G -> D (IN ALLELE 4).
SQ SEQUENCE 440 AA; 47118 MW; 4A1F7393 CRC32;

Query Match 36.4%; Score 64; DB 1; Length 440;
Best Local Similarity 35.3%; Pred. No. 6.88e+00;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 151 MAKILASGQVYISDAF 167
| | : | : | : |
QY 1 MEKFMAEFGQGYVQTPF 17

Search completed: Wed Sep 1 16:05:11 1999
Job time : 11 secs.

SEQUENCE FROM N.A.;
RC STRAIN=CV. COLUMBIA;
RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERES C.R., VENTER J.C.;

Query Match 38.6%; Score 68; DB 2; Length 418;
Best Local Similarity 33.3%; Pred. No. 3.70e+00;
Matches 7; Conservative 9; Mismatches 5; Indels

0: Gaps 0:

Db 177 FLGHFSGFVMTYLAQGLNV 197
I:: I:: I:: I:: I:: I::
QY 4 FMAEFGGVQTPFLSESNSV 24

RESULT 7
ID O87325 PRELIMINARY: PRT: 470 AA.
AC O87325;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE LECITHINASE.
GN PHL.
OS VIBRIO MIMICUS.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; VIBRIACEAE; VIBRIO.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33653;
RX MEDLINE: 98440110.
RA KONG I.S.;
RT "Cloning and identification of a phospholipase gene from *Vibrio*
mimicus".
RL BIOCHIM. BIOPHYS. ACTA 1394:85-89(1998).
DR EMBL: AF035162; G3746409; -;
DR PROSITE: PS01098; LIPASE_GDSL_SER; 1.
SQ SEQUENCE 470 AA; 53194 MW; 1EF1E5A5 CRC32;

Query Match 38.6%; Score 68; DB 2; Length 470;
Best Local Similarity 33.3%; Pred. No. 3.70e+00;
Matches 7; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

Db 229 FLGHFSGFVMTYLAQGLNV 249
I:: I:: I:: I:: I:: I::
QY 4 FMAEFGGVQTPFLSESNSV 24

RESULT 8
ID Q61138 PRELIMINARY: PRT: 1378 AA.
AC Q61138;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PATERNALLY EXPRESSED PROTEIN 3 (2N-FINGER PROTEIN PW1).
GN PEG3.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SV129; TISSUE-LIMB BUD;
RX MEDLINE: 96400442.
RA RELAIX F., WENG X., MARAZZI G., YANG E., COPELAND N., JENKINS N.,
RA SPENCE S.E., SASSOON D.;
RT "Pw1, a novel zinc finger gene implicated in the myogenic and
neuronal lineages.";
RL DEV. BIOL. 177:383-396(1996).
DR EMBL: G48804; G1197857; -;
DR MGD: MGI:104748; PEG3.
DR PROSITE: PS00028; ZINC_FINGER_C2H2; 11.
DR PFAM: PF00096; zf-C2H2; 11.
KW ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ SEQUENCE 1378 AA; 156916 MW; 756F3644 CRC32;

Query Match 38.6%; Score 68; DB 11; Length 1378;
Best Local Similarity 40.0%; Pred. No. 3.70e+00;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 1102 EYGPSTYTHASFLTEP 1116
I:: I:: I:: I:: I:: I::
QY 7 EFGQGVQTPFLSES 21

RESULT 9
ID O54978 PRELIMINARY: PRT: 1571 AA.
AC O54978;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ZINC FINGER PROTEIN.
GN PEG3.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96154192.
RA KUROIWA Y., KANEKO-ISHINO T., KAGITANI F., KOHDA T., LI L.L., TADA M.,
RA SUZUKI R., YOKOYAMA M., SHIROISHI T., WAKANA S., BARTON S.C.,
RA ISHINO F., SURANI M.A.;
RT "Peg3 imprinted gene on proximal chromosome 7 encodes for a zinc
finger protein.";
RL NAT. GENET. 12:186-190(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA SURANI A.M.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF038939; G2791678; -;
DR PROSITE: PS00028; ZINC_FINGER_C2H2; 11.
KW ZINC-FINGER; METAL-BINDING; DNA-BINDING.
SQ SEQUENCE 1571 AA; 178824 MW; 6F2820E6 CRC32;

Query Match 38.6%; Score 68; DB 11; Length 1571;
Best Local Similarity 40.0%; Pred. No. 3.70e+00;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 1295 EYGPSTYTHASFLTEP 1309
I:: I:: I:: I:: I:: I::
QY 7 EFGQGVQTPFLSES 21

RESULT 10
ID P75688 PRELIMINARY: PRT: 70 AA.
AC P75688;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE FROM BASES 311709 TO 323910
DE (SECTION 27 OF 400) OF THE COMPLETE GENOME (SECTION 27 OF 400).
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K-12;
RX MEDLINE: 97426617.
RA BLATTNER F.R., PLUNKETT III G., BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER J.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GORDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL SCIENCE 277:1453-1474(1997).
DR EMBL: AE000137; G1786500; -;
SQ SEQUENCE 70 AA; 7854 MW; B754D938 CRC32;

Query Match 37.5%; Score 66; DB 2; Length 70;
Best Local Similarity 53.3%; Pred. No. 7.55e+00;
Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 47 QYVQTSYLSQQSSI 61
I:: I:: I:: I:: I:: I::
QY 10 QYVQTPFLSESNSV 24

RESULT 11
ID O87765 PRELIMINARY: PRT: 215 AA.

AC 087765;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE PYROLIDONE CARBOXYL PEPTIDASE (EC 3.4.19.1)
 DE (ACYLAMINOACYL-PEPTIDASE) (ACYLAMINO-ACID RELEASING ENZYME)
 DE (N-ACYLPEPTIDE HYDROLASE)
 GN PCP
 OS LACTOCOCCUS LACTIS SUBSP. CREMORIS.
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC LACTOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 98406035.
 RA DAVERAN-MINGOT M.L., CAMPO N., RITZENTHALER P., LE BOURGEOIS P.;
 RT "A natural large chromosomal inversion in *Lactococcus lactis* is
 RT mediated by homologous recombination between two insertion
 RT sequences."
 RL J. BACTERIOL. 180:4834-4842(1998).
 CC -1- CATALYTIC ACTIVITY: ACYLAMINOACYL-PEPTIDE + H(2)O = ACYLAMINO
 CC ACID + PEPTIDE.
 DR EMBL: AJ2233960; E1323722;
 KW HYDROLASE.
 SQ SEQUENCE 215 AA; 23519 MW; 004FCDDF CRC32;
 Query Match 37.5%; Score 66; DB 2; Length 215;
 Best Local Similarity 31.6%; Pred. No. 7.55e+00;
 Matches 6; Conservative 8; Mismatches 5; Indels 0; Gaps 0;
 Db 153 DKMPDPTKAGPMHPPFLPE 171
 QY 2 EKFMAEFGQGYVOTPFLLSE 20
 RESULT 12
 ID Q47402 PRELIMINARY; PRT; 296 AA.
 AC Q47402;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE REGION 1 GENE PRODUCT (FRAGMENT).
 GN KPSS.
 OS ESCHERICHIA COLI.
 OC BACTERIA: PROTEOBACTERIA: GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
 OC ESCHERICHIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92138601.
 RA STEENBERGEN S.M., WRONA T.J., VIMR E.R.;
 RT "Functional analysis of the sialyltransferase complexes in
 RT *Escherichia coli* K1 and K92."
 RL J. BACTERIOL. 174:1099-1108(1992).
 DR EMBL: M76370; G146950;
 FT NON-TER
 SQ SEQUENCE 296 AA; 35733 MW; E10BA47B CRC32;
 Query Match 37.5%; Score 66; DB 2; Length 296;
 Best Local Similarity 34.8%; Pred. No. 7.55e+00;
 Matches 8; Conservative 9; Mismatches 4; Indels 2; Gaps 2;
 Db 2 RFLA-FREGYLRPOFITVEEDGV 23
 QY 3 KFMAEFGQGYVOTPFLLS-ESNSV 24
 RESULT 13
 ID Q23414 PRELIMINARY; PRT; 334 AA.
 AC Q23414;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE COSMID ZK1248.
 GN ZK1248.5.

OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA: METAZOA: NEMATODA: SECERNENTEA: RHABDITIA: RHABDITIDA;
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-BRISTOL N2;
 RX MEDLINE: 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SALLDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*
 RT *elegans*."
 RL NATURE 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA LATREILLE P.;
 RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: U29244; G862494;
 SQ SEQUENCE 334 AA; 38293 MW; 9ACC97C2 CRC32;
 Query Match 37.5%; Score 66; DB 5; Length 334;
 Best Local Similarity 39.1%; Pred. No. 7.55e+00;
 Matches 9; Conservative 4; Mismatches 9; Indels 1; Gaps 1;
 Db 55 IKKFKAWGNDYHOTLFLKDS 77
 QY 1 MEKFMAEFGQGYVOT-PFLSEN 22
 RESULT 14
 ID Q58468 PRELIMINARY; PRT; 336 AA.
 AC Q58468;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE 336AA LONG HYPOTHETICAL PROTEIN.
 GN PH0737.
 OS PYROCOCCLUS HORIKOSHII.
 OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCCLUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OT3;
 RX MEDLINE: 98344137.
 RA KAWABAYASHI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
 RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOTAMA A., NAGAI Y.,
 RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
 RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
 RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
 RA KIKUCHI H.;
 RT "Complete Sequence and Gene Organization of the Genome of a
 RT Hyper-thermophilic Archaeobacterium, *Pyrococcus horikoshii* OT3."
 RL DNA RES. 5:55-76(1998).
 DR EMBL: AF000003; D1030771;
 SQ SEQUENCE 336 AA; 37579 MW; 62430DEB CRC32;
 Query Match 37.5%; Score 66; DB 1; Length 336;
 Best Local Similarity 38.9%; Pred. No. 7.55e+00;
 Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
 Db 166 EYVNGFVKSHFLDDKASV 183

 W P S R L A
 (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:09:02 1999; MasPar time 3.86 Seconds
 Tabular output not generated. 93.552 Million cell updates/sec

Title: >PCT-US99-13024-2

Description: (1-17) from PCTUS9913024.pep (3 of 12)

Perfect Score: 138
 Sequence: 1 MEKFMAEFGQGVQPF 17

Scoring table: PAM 150
 Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33
 34:part34 35:part35 36:part36 37:part37 38:part38
 39:part39

Statistics: Mean 20.181; Variance 71.670; scale 0.282

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	63	45.7	389	2	D-alanine racemase.	4.48e+01
2	60	43.5	130	36	S. pneumoniae protein	8.42e+01
3	59	42.8	38	35	Leader peptide sequen	1.04e+02
4	59	42.8	364	33	Streptococcus pneumon	1.04e+02
5	59	42.8	380	35	Glycerol dehydrogenas	1.04e+02
6	59	42.8	409	9	Sequence of protease	1.04e+02
7	59	42.8	740	33	Streptococcus pneumon	1.04e+02
8	58	42.0	578	23	Human oxalyl-CoA deca	1.28e+02
9	58	42.0	578	14	Oxalyl-CoA decarboxyl	1.28e+02
10	58	42.0	2721	2	Sequence of N-termin	1.28e+02
11	58	42.0	4536	28	Apolipoprotein B-100.	1.28e+02
12	58	42.0	5069	30	A. mediterranei rifam	1.28e+02
13	57	41.3	65	19	Phage T4 ORFX gene pr	1.57e+02
14	57	41.3	321	8	Carnitine dehydrogena	1.57e+02
15	57	41.3	623	36	HSV-2 strain SB5 Cont	1.57e+02
16	57	41.3	1196	36	HSV-2 strain SB5 Cont	1.57e+02

17	57	41.3	1670	28	W41314	Full length heat-resi	1.57e+02
18	57	41.3 <td>1670 <td>16 <td>R85599 <td>DNA polymerase from s <td>1.57e+02</td> </td></td></td></td>	1670 <td>16 <td>R85599 <td>DNA polymerase from s <td>1.57e+02</td> </td></td></td>	16 <td>R85599 <td>DNA polymerase from s <td>1.57e+02</td> </td></td>	R85599 <td>DNA polymerase from s <td>1.57e+02</td> </td>	DNA polymerase from s <td>1.57e+02</td>	1.57e+02
19	57	41.3 <td>1670 <td>18 <td>R94616 <td>DNA polymerase. <td>1.57e+02</td> </td></td></td></td>	1670 <td>18 <td>R94616 <td>DNA polymerase. <td>1.57e+02</td> </td></td></td>	18 <td>R94616 <td>DNA polymerase. <td>1.57e+02</td> </td></td>	R94616 <td>DNA polymerase. <td>1.57e+02</td> </td>	DNA polymerase. <td>1.57e+02</td>	1.57e+02
20	57	41.3 <td>1670 <td>18 <td>R97048 <td>DNA polymerase, contg <td>1.57e+02</td> </td></td></td></td>	1670 <td>18 <td>R97048 <td>DNA polymerase, contg <td>1.57e+02</td> </td></td></td>	18 <td>R97048 <td>DNA polymerase, contg <td>1.57e+02</td> </td></td>	R97048 <td>DNA polymerase, contg <td>1.57e+02</td> </td>	DNA polymerase, contg <td>1.57e+02</td>	1.57e+02
21	57	41.3 <td>1670 <td>23 <td>W20049 <td>KOD1 thermostable DNA <td>1.57e+02</td> </td></td></td></td>	1670 <td>23 <td>W20049 <td>KOD1 thermostable DNA <td>1.57e+02</td> </td></td></td>	23 <td>W20049 <td>KOD1 thermostable DNA <td>1.57e+02</td> </td></td>	W20049 <td>KOD1 thermostable DNA <td>1.57e+02</td> </td>	KOD1 thermostable DNA <td>1.57e+02</td>	1.57e+02
22	56	40.6 <td>43</td> <td>8 <td>R40095 <td>Hib OMP Pl-P2 hybrid <td>1.92e+02</td> </td></td></td>	43	8 <td>R40095 <td>Hib OMP Pl-P2 hybrid <td>1.92e+02</td> </td></td>	R40095 <td>Hib OMP Pl-P2 hybrid <td>1.92e+02</td> </td>	Hib OMP Pl-P2 hybrid <td>1.92e+02</td>	1.92e+02
23	56	40.6 <td>461</td> <td>30 <td>W39928 <td>Human CDP-diacyglyce <td>1.92e+02</td> </td></td></td>	461	30 <td>W39928 <td>Human CDP-diacyglyce <td>1.92e+02</td> </td></td>	W39928 <td>Human CDP-diacyglyce <td>1.92e+02</td> </td>	Human CDP-diacyglyce <td>1.92e+02</td>	1.92e+02
24	56	40.6 <td>979</td> <td>39 <td>W83068 <td>TuT protein of tolu <td>1.92e+02</td> </td></td></td>	979	39 <td>W83068 <td>TuT protein of tolu <td>1.92e+02</td> </td></td>	W83068 <td>TuT protein of tolu <td>1.92e+02</td> </td>	TuT protein of tolu <td>1.92e+02</td>	1.92e+02
25	55	39.9	17	12 <td>R63065 <td>Acetyl-esterase II N- <td>2.36e+02</td> </td></td>	R63065 <td>Acetyl-esterase II N- <td>2.36e+02</td> </td>	Acetyl-esterase II N- <td>2.36e+02</td>	2.36e+02
26	55	39.9	218	39 <td>W89435 <td>Streptococcus pneumon <td>2.36e+02</td> </td></td>	W89435 <td>Streptococcus pneumon <td>2.36e+02</td> </td>	Streptococcus pneumon <td>2.36e+02</td>	2.36e+02
27	55	39.9	218	39 <td>W83371 <td>Streptococcus pneumon <td>2.36e+02</td> </td></td>	W83371 <td>Streptococcus pneumon <td>2.36e+02</td> </td>	Streptococcus pneumon <td>2.36e+02</td>	2.36e+02
28	55	39.9	486	22 <td>W20586 <td>H. pylori cytoplasmic <td>2.36e+02</td> </td></td>	W20586 <td>H. pylori cytoplasmic <td>2.36e+02</td> </td>	H. pylori cytoplasmic <td>2.36e+02</td>	2.36e+02
29	55	39.9	518	39 <td>W67616 <td>A. nidulans phenylace <td>2.36e+02</td> </td></td>	W67616 <td>A. nidulans phenylace <td>2.36e+02</td> </td>	A. nidulans phenylace <td>2.36e+02</td>	2.36e+02
30	55	39.9	622	8 <td>R38888 <td>Sequence encoded by O <td>2.36e+02</td> </td></td>	R38888 <td>Sequence encoded by O <td>2.36e+02</td> </td>	Sequence encoded by O <td>2.36e+02</td>	2.36e+02
31	54	39.1	23	10 <td>R53794 <td>T. litoralis IVPS1 (3 <td>2.89e+02</td> </td></td>	R53794 <td>T. litoralis IVPS1 (3 <td>2.89e+02</td> </td>	T. litoralis IVPS1 (3 <td>2.89e+02</td>	2.89e+02
32	54	39.1	23	37 <td>W70570 <td>Protein-splice juncti <td>2.89e+02</td> </td></td>	W70570 <td>Protein-splice juncti <td>2.89e+02</td> </td>	Protein-splice juncti <td>2.89e+02</td>	2.89e+02
33	54	39.1	344	8 <td>R40918 <td>Bacillus subtilis ino <td>2.89e+02</td> </td></td>	R40918 <td>Bacillus subtilis ino <td>2.89e+02</td> </td>	Bacillus subtilis ino <td>2.89e+02</td>	2.89e+02
34	54	39.1	458	4 <td>R22279 <td>Human gp.-specific co <td>2.89e+02</td> </td></td>	R22279 <td>Human gp.-specific co <td>2.89e+02</td> </td>	Human gp.-specific co <td>2.89e+02</td>	2.89e+02
35	54	39.1	458	4 <td>R22278 <td>Human gp.-specific co <td>2.89e+02</td> </td></td>	R22278 <td>Human gp.-specific co <td>2.89e+02</td> </td>	Human gp.-specific co <td>2.89e+02</td>	2.89e+02
36	54	39.1	458	23 <td>W10561 <td>Vitamin D3-binding pr <td>2.89e+02</td> </td></td>	W10561 <td>Vitamin D3-binding pr <td>2.89e+02</td> </td>	Vitamin D3-binding pr <td>2.89e+02</td>	2.89e+02
37	54	39.1	458	11 <td>R56975 <td>Variant vitamin D bin <td>2.89e+02</td> </td></td>	R56975 <td>Variant vitamin D bin <td>2.89e+02</td> </td>	Variant vitamin D bin <td>2.89e+02</td>	2.89e+02
38	54	39.1	481	26 <td>W34554 <td>M17L glycosidase 29G <td>2.89e+02</td> </td></td>	W34554 <td>M17L glycosidase 29G <td>2.89e+02</td> </td>	M17L glycosidase 29G <td>2.89e+02</td>	2.89e+02
39	54	39.1	621	29 <td>W55645 <td>H. pylori ORF 06ep106 <td>2.89e+02</td> </td></td>	W55645 <td>H. pylori ORF 06ep106 <td>2.89e+02</td> </td>	H. pylori ORF 06ep106 <td>2.89e+02</td>	2.89e+02
40	54	39.1	645	37 <td>W73030 <td>Helicobacter pylori 7 <td>2.89e+02</td> </td></td>	W73030 <td>Helicobacter pylori 7 <td>2.89e+02</td> </td>	Helicobacter pylori 7 <td>2.89e+02</td>	2.89e+02
41	54	39.1	834	2 <td>R08391 <td>Sequence encoded by v <td>2.89e+02</td> </td></td>	R08391 <td>Sequence encoded by v <td>2.89e+02</td> </td>	Sequence encoded by v <td>2.89e+02</td>	2.89e+02
42	54	39.1	911	18 <td>R95634 <td>DNA-ligase-III. <td>2.89e+02</td> </td></td>	R95634 <td>DNA-ligase-III. <td>2.89e+02</td> </td>	DNA-ligase-III. <td>2.89e+02</td>	2.89e+02
43	54	39.1	1312	10 <td>R53787 <td>Deep Vent DNA polymer <td>2.89e+02</td> </td></td>	R53787 <td>Deep Vent DNA polymer <td>2.89e+02</td> </td>	Deep Vent DNA polymer <td>2.89e+02</td>	2.89e+02
44	54	39.1	1312	37 <td>W70563 <td>DNA polymerase of a <td>2.89e+02</td> </td></td>	W70563 <td>DNA polymerase of a <td>2.89e+02</td> </td>	DNA polymerase of a <td>2.89e+02</td>	2.89e+02
45	54	39.1	1829	27 <td>W29322 <td>DNA polymerase with 3 <td>2.89e+02</td> </td></td>	W29322 <td>DNA polymerase with 3 <td>2.89e+02</td> </td>	DNA polymerase with 3 <td>2.89e+02</td>	2.89e+02

ALIGNMENTS

RESULT 1
 ID P70668 standard; protein; 389 AA.
 AC P70668;
 DT 11-MAR-1991 (first entry)
 DE D-alanine racemase.
 KW L-glutamate racemase; D-alanyl-D-alanine ligase.
 OS Bacillus subtilis.
 PN GB2177097-A.
 PD 14-JAN-1987.
 PF 17-JUN-1986; 014702.
 PR 18-JUN-1985; US-746437.
 PA (GENE-) GENENCOR INC.
 PI Ferrari E;
 DR WPI; 87-009482/02.
 DR N-PSDB; N70668.
 PT Stable maintenance of heterologous DNA in host cell - using host
 PT cells deficient in ability to synthesise cell wall transformed to
 PT regain ability
 PS Disclosure; Fig 2; 13pp; English.
 CC Protein may be produced in a transformed cell-line deficient in
 CC D-alanine production. The cell will then express the required sequence
 CC and a second sequence operably linked to the D-alanine racemase.
 CC Other sequences which may be used in the same way include those
 CC encoding L-glutamate racemase or D-alanyl-D-alanine ligase.
 SQ Sequence 389 AA;
 Query Match 45.7%; Score 63; DB 2; Length 389;
 Best Local Similarity 50.0%; Pred. No. 4.48e+01;
 Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 Db 313 mdqfmevdqey 324
 QY I:|:|:|:|:
 1 MEKFMAEFGQGV 12
 RESULT 2
 ID W80717 standard; Protein; 130 AA.
 AC W80717;
 DT 24-DEC-1998 (first entry)
 DE S. pneumoniae protein of unknown function.

KW Streptococcus pneumoniae protein; recombinant; gene expression; DNA chip;
 KW virulence; antibody; infection; detection; treatment; hypothetical;
 OS cell wall biosynthetic, external target; minimal gene set protein.
 PN Streptococcus pneumoniae.

PD WO9826072-A1.
 PF 18-JUN-1998.
 PR 09-DEC-1997; U22578.
 PR 13-DEC-1996; US-036281.
 PA (ELIL) LILLY & CO ELI.
 PI Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR,
 PI Mills BJ, Norris FH, Peery RB, Rockey PK, Rostack PR,
 PI Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ,
 PI Young Beilido ML;
 DR WPI; 98-348529/30.
 DR N-PSDB; V65274.

PT Streptococcus pneumoniae nucleic acid sequences - used in DNA chips
 PT for evaluating gene expression, and identification of virulence

PT genes
 PS Claim 3: Pages 306-307; 333pp; English.
 CC This sequence represents a Streptococcus pneumoniae protein of unknown
 CC function. The invention provides DNA sequences (V65201 to V65304) from
 CC the Streptococcus pneumoniae genome and corresponding protein sequences
 CC (W60605 to W80728). The protein sequences are classified as hypothetical,
 CC cell wall biosynthetic, external target, or minimal gene set proteins. A
 CC recombinant host containing a vector comprising any of the above nucleic
 CC acids can be used for the recombinant expression of the proteins. The
 CC invention also provides a DNA chip having arrayed on it at least 15 base
 CC pair fragment of any one or more of these DNA sequences. The DNA chip can
 CC be used methods for evaluating gene expression in S. pneumoniae and for
 CC identifying virulence genes in S. pneumoniae. Antibodies that selectively
 CC bind to the above proteins or peptide fragments can be used to treat
 CC S. pneumoniae infection. The antibodies can also be used to detect
 CC S. pneumoniae cells.
 CC Sequence 130 AA;

Query Match 43.5%; Score 60; DB 36; Length 130;
 Best Local Similarity 53.8%; Pred. No. 8.42e+01;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 69 eqimadlaqvqvq 81
 QY I : : : : : I I I
 2 EKFMAEFGQGYVQ 14

RESULT 3
 ID W8736 standard; peptide: 38 AA.

AC W8736;
 DT 05-NOV-1998 (first entry)
 DE Leader peptide sequence TA75 50.
 KW YAP3 signal peptide; synthetic leader peptide; construct; yeast;
 KW insulin; insulin-like growth factor; glucagon.
 OS Synthetic.
 PN WO9832867-A1.
 PD 30-JUL-1998.
 PR 22-JAN-1998; DK00026.
 PR 24-JAN-1997; DK-000097.
 PA (NOVO) NOVO-NORDISK AS.
 PI Balschmidt P, Havelund S, Kjellden TB, Pettersson AF;
 DR WPI; 98-427963/36.

PT New DNA construct containing synthetic leader sequence with no
 PT N-linked glycosylation - for high yield and efficient expression and
 PT secretion of proteins, e.g. insulin, in yeast, also related vectors
 PT and transformed yeast
 PS Claim 14; Page 10; 41pp; English.
 CC A DNA construct has been developed encoding a polypeptide and has the
 CC formula given below, components in parentheses may be omitted:
 CC SP-LP-(PS)-(S)-(PS)-gene (I); where SP = sequence encoding a signal
 CC peptide; LP = sequence encoding synthetic leader peptide without any
 CC N-linked glycosylation; PS = sequence encoding a protease processing
 CC site; S = sequence encoding a spacer peptide; gene = sequence encoding a
 CC polypeptide. Also described are: (1) an expression cassette (EC)
 CC comprising (I), fused to a 5'-promoter and 3'-terminator; (2) a yeast
 CC expression vector containing (I); (3) a yeast cell containing the vector

CC as in (2); (4) DNA encoding a synthetic preproleader sequence that lacks
 CC the consensus N-linked glycosylation site NXT/S (X = any codable amino
 CC acid other than P); and (5) the preproleader sequence. Expression vectors
 CC of (2) are used to produce (both express and secrete) polypeptides in
 CC yeast, e.g. insulin, insulin-like growth factor, glucagon or its
 CC fragments. The synthetic preproleader sequences are used to direct
 CC secretion of proteins from yeast cells. These LP provide high yield and
 CC more efficient recovery and/or purification of secreted proteins
 CC expressed in eukaryotic cells. The present sequence represents a
 CC specifically claimed leader peptide sequence of the present invention.
 CC Sequence 38 AA;

Query Match 42.8%; Score 59; DB 35; Length 38;
 Best Local Similarity 35.3%; Pred. No. 1.04e+02;
 Matches 6; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

Db 11 mvnlmadddgfmqapl 27
 QY I : : : : : I : : : : : I
 1 MEKFMAEFGQGYVQTPF 17

RESULT 4

ID W62047 standard; Protein: 364 AA.
 AC W62047;
 DT 01-OCT-1998 (first entry)
 DE Streptococcus pneumoniae spo/rei distal terminal portion.
 KW Streptococcus pneumoniae; spo/rei; spo/reiA family; screening;
 KW antibacterial; otitis media; conjunctivitis; pneumonia; bacteraemia;
 KW meningitis.
 OS Streptococcus pneumoniae.
 PN EP-849362-A2.
 PD 24-JUN-1998.
 PR 23-OCT-1997; 308473.
 PR 24-OCT-1996; US-029049.
 PA (SMIK) SMITHKLINE BEECHAM CORP.

PI Gentry DR;
 DR N-PSDB; V37948.
 DR New Streptococcus spo/reiA polypeptide(s) - useful in treatment of
 PT otitis media, conjunctivitis, pneumonia, bacteraemia and especially
 PT meningitis
 PS Claim 11; Page 31-32; 40pp; English.
 CC The present sequence is the distal terminal portion of a spo/rei
 CC protein (spo/reiA family) from Streptococcus pneumoniae. The protein is
 CC of the spo/reiA family, which, in Escherichia coli, is involved in the
 CC stringent response to nutrient limitation and regulate the accumulation
 CC of (p)ppGpp which is involved in the regulation of gene expression and
 CC other cellular processes. Spo/rei proteins have a homology to e.g.
 CC S. equisimilis rel protein. Spo/rei proteins can be used: (a) in the
 CC treatment of an individual in need of spo/rei protein, by administering
 CC to the protein to the patient; (b) in the treatment of an individual
 CC having need to inhibit spo/rei protein, by administering an antagonist
 CC which inhibits the activity of the protein; and (c) for inducing an
 CC immunological response by inoculating the mammal with spo/rei protein,
 CC or a fragment or variant of it, adequate to produce antibody and/or T
 CC cell immune response to protect the animal from disease. Conditions
 CC which may be treated include otitis media, conjunctivitis, pneumonia,
 CC bacteraemia, meningitis, sinusitis, pleural empyema, endocarditis and
 CC especially meningitis and compositions may be used as antibacterials.
 CC Sequence 364 AA;

Query Match 42.8%; Score 59; DB 33; Length 364;
 Best Local Similarity 47.1%; Pred. No. 1.04e+02;
 Matches 8; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

Db 119 emlmaqfnggyvankf 135
 QY I : : : : : I : : : : : I
 2 EKFMAEFGQ-GYVQTPF 17

RESULT 5

ID W69360 standard; Protein: 380 AA.
 AC W69360;

DT 01-DEC-1998 (first entry)
 DE Glycerol dehydrogenase.
 KW 1,2-propanediol production; intracellular methylglyoxal conversion;
 KW unsaturated polyester resin; liquid laundry detergent; antifreeze;
 KW de-icing composition; animal feed; glycerol dehydrogenase.
 OS Escherichia coli.
 FH Key Location/Qualifiers
 FT Misc_difference 42
 FT /note= "unspecified amino acid"
 PN WO9837204-A1.
 PD 27-AUG-1998.
 PF 19-FEB-1998; U03271.
 PR 19-FEB-1997; US-801344.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 PI Altaras NE, Cameron DC, Shaw AJ;
 DR WPI; 98-480799/41.
 PT Fermentative production of 1,2-propanediol from simple sugars -
 PT using recombinant microorganism that expresses enzymes that convert
 PT extracellular methylglyoxal, also new transformants and synthetic
 PT operons
 PS Disclosure; Page 42-43; 59pp; English.
 CC This sequence is the E. coli glycerol dehydrogenase. The DNA
 CC can be used in a vector used in the method of the invention. The
 CC method is for the production of 1,2-propanediol (1), and comprises
 CC culturing a recombinant microorganism (A) that expresses enzymes(s) able
 CC to convert intracellular methylglyoxal (MG) to (1), in a medium
 CC containing a sugar, other than a 6-deoxyhexose, as carbon source. (1) is
 CC used in manufacture of unsaturated polyester resins, liquid laundry
 CC detergents, pharmaceuticals, cosmetics, antifreeze and de-icing
 CC compositions. By-products of the process (carbon dioxide, alcohols and
 CC organic acids) may be used e.g. as animal feeds. The method uses
 CC inexpensive sugars and provides production of (1) from renewable sources,
 CC without: (a) use of propylene oxide as intermediate; (b) generation of
 CC toxic waste; and (c) use of high temperature and pressure. Yields of (1)
 CC of up to 100 g/l can be achieved.
 CC Sequence 380 AA;
 SQ

 Query Match 42.8%; Score 59; DB 35; Length 380;
 Best Local Similarity 37.5%; Pred. No. 1.04e+02;
 Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

 Db 50 dkfvlfgagstveksf 65
 QY 2 EKFMFEFGGQYVQTPF 17
 :||: ||:| |:

 RESULT 6
 ID R48060 standard; Protein; 409 AA.
 AC R48060.
 DE 20-JUL-1994 (first entry)
 DT Sequence of protease A encoded by gene PRAL.
 KW Protease; yeast; proteolysis.
 OS Kluyveromyces fragilis.
 PN WO9400579-A.
 PD 06-JAN-1994.
 PF 23-JUN-1993; F00623.
 PR 25-JUN-1992; FR-007785.
 PA (RHON) RHONE POULENC RORER SA.
 PI Fleer R., Fournier A, Yeh P;
 DR WPI; 94-026215/03.
 DR N-PSDB; Q55348.
 PT New Kluyveromyces yeast with modified protease gene - esp. used
 PT for high yield prodn. of recombinant protein, also DNA encoding
 PT yeast protease and derived peptide(s)
 PS Disclosure; Page 31-33; 49pp; English.
 CC The protease gene is to be modified in order to render it (partially)
 CC incapable of producing the natural protein; or result in a non-
 CC functional protease or in a protease with modified proteolytic
 CC activity. The modifications can be introduced in vitro or in situ by
 CC standard genetic engineering techniques or by exposure to mutagenic
 CC agents.
 CC Sequence 409 AA;
 SQ

Query Match 42.8%; Score 59; DB 9; Length 409;
 Best Local Similarity 37.5%; Pred. No. 1.04e+02;
 Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

 Db 54 ekylnsfqrqaypqs 69
 QY 2 EKFMFEFGGQYVQTPF 17
 ||:| |:

 RESULT 7
 ID W62045 standard; Protein; 740 AA.
 AC W62045;
 DT 01-OCT-1998 (first entry)
 DE Streptococcus pneumoniae spo/rel protein sequence.
 KW Streptococcus pneumoniae; spo/rel; spo/relA family; screening;
 KW antibacterial; otitis media; conjunctivitis; pneumonia; bacteraemia;
 KW meningitis.
 OS Streptococcus pneumoniae.
 PN EP-849362-A2.
 PD 24-JUN-1998.
 PF 23-OCT-1997; 308473.
 PR 24-OCT-1996; US-029049.
 PA (SMIK) SMITHLINE BEECHAM CORP.
 PI Gentry DR;
 DR WPI; 98-324673/29.
 DR N-PSDB; V37946.
 PT New Streptococcus spo/relA polypeptide(s) - useful in treatment of
 PT otitis media, conjunctivitis, pneumonia, bacteraemia and especially
 PT meningitis
 PS Claim 1; Page 26-28; 40pp; English.
 CC The present sequence represents a spo/rel protein (spo/relA family) from
 CC Streptococcus pneumoniae. The protein is of the spo/relA family, which,
 CC in Escherichia coli, is involved in the stringent response to nutrient
 CC limitation and regulate the accumulation of (p)ppGpp which is involved
 CC in the regulation of gene expression and other cellular processes.
 CC Spo/rel proteins have a homology to e.g. S. equisimilis rel protein.
 CC Spo/rel proteins can be used: (a) in the treatment of an individual in
 CC need of spo/rel protein, by administering to the protein to the patient;
 CC (B) in the treatment of an individual having need to inhibit spo/rel
 CC protein, by administering an antagonist which inhibits the activity of
 CC the protein; and (C) for inducing an immunological response by
 CC inoculating the mammal with spo/rel protein, or a fragment or variant
 CC of it, adequate to produce antibody and/or T cell immune response to
 CC protect the animal from disease. Conditions which may be treated
 CC include otitis media, conjunctivitis, pneumonia, bacteraemia, meningitis,
 CC sinusitis, pleural empyema, endocarditis and especially meningitis and
 CC compositions may be used as antibacterials.
 CC Sequence 740 AA;
 SQ

 Query Match 42.8%; Score 59; DB 33; Length 740;
 Best Local Similarity 47.1%; Pred. No. 1.04e+02;
 Matches 8; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

 Db 495 emlmaqfnggyvankf 511
 QY 2 EKFMFEFGQ-YVQTPF 17
 |::||:| |:

 RESULT 8
 ID W22882 standard; Protein; 578 AA.
 AC W22882;
 DT 24-SEP-1997 (first entry)
 DE Human oxalyl-CoA decarboxylase.
 KW Human; oxalyl-CoA; decarboxylase; prevention; hyperoxaluria;
 KW urinary stone; oxalate ion; detection; diagnosis; treatment.
 OS Homo sapiens.
 PN US5635616-A.
 PD 03-JUN-1997.
 PR 02-JUN-1995; 458120.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (GENO-) INST GENOMIC RES.
 PI Adams MD, Coleman TA, Olsen HS;

```

DR WPI: 97-309885/28.
DR N-PSDB: T75517.
PT Isolated oxalyl-CoA decarboxylase polynucleotide(s) - useful to
PT produce enzymes, which can be used to prevent urinary stone
PT formation or treat or prevent hyperoxaluria
PS Claim 1: Columns 17-22: 20pp; English.
CC The present sequence is human oxalyl-CoA decarboxylase, which
CC can be used to prevent or treat hyperoxaluria or urinary stone
CC formation by reducing the plasma or urinary levels of oxalate ion.
CC The enzyme or the polynucleotide encoding it can also be used for
CC detection and diagnosis.
CC The polynucleotide was identified in a cDNA library derived from
CC the human pancreas. It contains an open reading frame encoding
CC the mature protein, which is about 50-60% homologous to the
CC oxalyl-CoA decarboxylase from the bacterium Oxalobacter formigenes.
SQ Sequence 578 AA;

Query Match 42.0%; Score 58; DB 23; Length 578;
Best Local Similarity 58.8%; Pred. No. 1.28e+02;
Matches 10; Conservative 1; Mismatches 4; Indels 2; Gaps 2;

Db 519 eqvmtafgkggyfvgtg 535
1 : 11 11111
2 EKFMFEFG-QGY-VQTP 16

QY 2 EKFMFEFG-QGY-VQTP 16

RESULT 9
ID R75085 standard; Protein: 578 AA.
AC R75085;
DT 23-JAN-1996 (first entry)
DE Oxalyl-CoA decarboxylase.
KW Oxalyl-CoA decarboxylase; hyperoxaluria; gene therapy;
KW fluorescent in situ hybridisation: FISH.
OS Homo sapiens.
FH Key Location/Qualifiers
FT region 1..517
FT /note= "this region is given in fig 1, however, the
FT mature protein is 578 amino acids in length".
PN ZA9403789-A.
PD 26-APR-1995.
PR 30-MAY-1994; 003789.
PR 30-MAY-1994; ZA-003789.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Adams MD, Coleman TA, Olsen H;
DR WPI: 95-207157/27.
DR N-PSDB: Q90522.
PT Isolated polynucleotide encoding oxalyl-CoA decarboxylase - useful
PT for protein prodn. for treating hyper:oxaluria
PS Claim 5; Page 34-35; 50pp; English.
CC This sequence represents human oxalyl-CoA decarboxylase. Oxalyl-CoA
CC decarboxylase may be used to prevent urinary stone formation by reducing
CC the plasma or urinary levels of the oxalate ion. It may also be used to
CC prevent hyperoxaluria. Hyperoxaluria is characterised by either abnormal
CC synthesis or hyper-absorption of oxalate which can be prevented by
CC degrading the oxalate ions. This sequence can also be used to identify
CC other molecules with similar biological activity, and may be used as an
CC immunogen to produce specific antibodies. The cDNA that encodes this
CC sequence may be used in gene therapy and also for chromosome
CC identification, e.g. fluorescent in situ hybridisation (FISH) of a cDNA
CC clone to a metaphase chromosomal spread.
SQ Sequence 578 AA;

Query Match 42.0%; Score 58; DB 14; Length 578;
Best Local Similarity 58.8%; Pred. No. 1.28e+02;
Matches 10; Conservative 1; Mismatches 4; Indels 2; Gaps 2;

Db 519 eqvmtafgkggyfvgtg 535
1 : 11 11111
2 EKFMFEFG-QGY-VQTP 16

QY 2 EKFMFEFG-QGY-VQTP 16

RESULT 10
ID P70647 standard; protein: 2721 AA.

```

```

AC P70647;
DT 17-APR-1991 (first entry)
DE Sequence of N-terminal apolipoprotein B (apoB).
KW Lipid-binding peptide; apolipoprotein purification.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 87..167
FT /label= Signal
FT 168..8247
PN W08702062-A.
PD 09-APR-1987.
PF 02-OCT-1986; U02075.
PR 04-OCT-1985; US-784418.
PR 04-DEC-1985; US-804692.
PR 27-FEB-1986; US-834300.
PA (BIOT-) BIOTECHN RES PARTN.
PI Protter AA, Vigne JL, Mallory JB, Talmadge KD;
DR WPI: 87-108703/15.
DR N-PSDB: N70996.
PT Purified lipid-binding peptide prodn. - by expression of
PT recombinant system and adding lipid to obtd. peptide mixt. to
PT form lipo-peptide complex
PS Claim 50; pp73-85; 123pp; English.
CC ApoB is a lipid-binding peptide adapted to bind to phospholipids at
CC one or more amphipatic alpha-helical peptide regions. ApoB gene is
CC pref. expressed in Chinese hamster ovary (CHO) cells and the gene is
CC introduced in an expression vector having a regulatable promoter
CC derived from the human metallothionein II gene.
SQ Sequence 2721 AA;

Query Match 42.0%; Score 58; DB 2; Length 2721;
Best Local Similarity 23.5%; Pred. No. 1.28e+02;
Matches 4; Conservative 7; Mismatches 5; Indels 1; Gaps 1;

Db 2174 lqymiqfdq-ykdsy 2189
1 : 11 11111
1 MEKFMFEFGQGYVQTPF 17

QY 1 MEKFMFEFGQGYVQTPF 17

RESULT 11
ID W41262 standard; peptide; 4536 AA.
AC W41262;
DT 19-MAY-1998 (first entry)
DE Apolipoprotein B-100.
KW Anti-coagulant; apolipoprotein B-100; apoB-100; metastatic spread;
KW thromboplasmin-mediated process; cancer; inhibitor; blood coagulation;
KW angiogenesis; cellular differentiation; apoptosis; KRAD-14;
KW prothrombinase complex.
OS Homo sapiens.
PN W09743311-A1.
PD 20-NOV-1997.
PR 09-MAY-1997; G01255.
PR 09-MAY-1996; GB-009702.
PA (UNLO) ROYAL FREE HOSPITAL SCHOOL MED.
PI Bruckdorfer KR, Ettelale C;
DR WPI: 98-008798/01.
PT Peptide fragments of apo:liipo:protein B-100 with anticoagulant
PT activity - used for treating or preventing coagulation, inhibiting
PT angiogenesis, cell differentiation and apoptosis
PS Disclosure; Page 42-47; 60pp; English.
CC This sequence represents the human apolipoprotein B-100 (apoB-100).
CC Fragments of this sequence can be used in the peptide of the invention,
CC which has the formula (I), or their variants with one or more internal
CC deletions, insertions or substitutions, while retaining anti-coagulant
CC properties of apolipoprotein B-100 (apoB-100).
CC Z1-KAQ-XI-KKKKKRHS-X2-T-22 (I); X1 = S or Y; X2 = T or I;
CC Z1 = the N terminus of the peptide, or 1-47 amino acids (aa);
CC Z2 = the C terminus of the peptide, a terminal amide group or 1-77 aa.
CC Compositions containing the peptide are used for simultaneous, separate
CC or sequential treatment of cancer, particularly to prevent metastatic
CC spread. They are also used to inhibit thromboplasmin-mediated processes,
CC specifically to prevent or reduce blood coagulation (e.g. during or after
CC surgery or in cases of heart attack, stroke etc.) and to inhibit

```

CC angiogenesis, cellular differentiation or apoptosis. KRAD-14, which is
CC active as such or as part of a 98-aa peptide, inhibits activation of the
CC prothrombinase complex; and prevents activation of factor VII on the
CC surface of thromboplastin and of platelets by thrombin. It binds to the
CC residues 58-66 of thromboplastin. Since (1) are much smaller than
CC apoB-100, they act more quickly.
SQ Sequence 4536 AA;

Query Match 42.0%; Score 58; DB 28; Length 4536;
Best Local Similarity 23.5%; Pred. No. 1.28e+02;
Matches 4; Conservative 7; Mismatches 5; Indels 1; Gaps 1;

Db 2147 lqymiqfdq-yikdsy 2162
::: |||||:::
QY 1 MEKFMFEFGQGYVQTPF 17

RESULT 12

ID W52846 standard; Protein; 5069 AA.
AC W52846;
DT 24-JUL-1998 (first entry)
DE A. mediterranei rifamycin synthesis gene cluster fragment protein B.
KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
OS polyketide synthase; actinomycete; ansamycin.
PN Amycolatopsis mediterranei.
PD 9807868-A1.
PD 26-FEB-1998.
PF 18-AUG-1997; E04495.
PR 20-AUG-1996; EP-810551.
PA (NOVS) NOVARTIS AG.
PI Engel N, Schupp T, Toupet C;
DR WPI; 98-169172/15.
DR N-PSDB; V21187.

PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT to produce rifamycin and rifamycin analogues
PS Claim 6; Page 126-151; 205pp; English.
CC The present sequence represents a Amycolatopsis mediterranei rifamycin
CC synthesis gene cluster ORF B protein from the present invention. The
CC DNA fragment comprises a DNA region involved directly or indirectly
CC in the gene cluster responsible for rifamycin synthesis, including
CC the adjacent DNA regions to the right and left which, by reason of
CC their function in connection with rifamycin biosynthesis, qualify
CC as constituents of this rifamycin gene cluster, and functional
CC fragments, derivatives or constituents of these. The Amycolatopsis
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC for producing rifamycin, rifamycin analogues or precursors. It can also
CC be used for inactivating or modifying genes involved in ansamycin or
CC rifamycin biosynthesis. The DNA can be used for constructing mutant
CC actinomycetes strains from which the natural rifamycin or ansamycin
CC biosynthesis gene cluster has been partly or completely deleted. The
CC DNA fragment can be used for assembling a library of polyketide
CC synthases, which can be used for assembling a library of polyketide
CC A hybridisation probe of the invention can be used for identifying DNA
CC fragments involved in the biosynthesis of ansamycins.
SQ Sequence 5069 AA;

Query Match 42.0%; Score 58; DB 30; Length 5069;
Best Local Similarity 70.0%; Pred. No. 1.28e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 3451 ftgmfgggv 3460
|:|||||
QY 4 FMAEFGQGYV 13

RESULT 13

ID R97371 standard; Protein; 65 AA.
AC R97371;
DT 07-JAN-1997 (first entry)
DE Phase T4 ORFX gene product. gp34.
KW Phase T4; tail fibre protein; nanotechnology; nano-structure;
KW filter; molecular sieve.
OS Bacteriophage T4.

PN WO9611947-A1.
PD 25-APR-1996.
PF 13-OCT-1995; U13023.
PR 13-OCT-1994; US-322760.
PA (GOLD/) GOLDBERG E B.
PI Goldberg EB;
DR WPI; 96-221942/22.
DR N-PSDB; T29053.

PT New proteins derived from T4 phage tail fibre proteins - that can
PT self assemble into nano-structure(s), useful as filters etc, also
PT corresponding DNA

PS Claim 7; Fig 7; 83pp; English.
CC A protein (R97371) of unspecified function is the product of open
CC reading frame X of the tail fibre protein gene region (see also
CC T29053) of phage T4. This gene region also includes open reading
CC frames for tail fibre proteins (see also R97370 and R97372-74).
CC Tail fibre proteins (native or modified) can be produced in large
CC quantities in microbial cells and used as building blocks of strong,
CC stable nanostructures.
SQ Sequence 65 AA;

Query Match 41.3%; Score 57; DB 19; Length 65;
Best Local Similarity 87.5%; Pred. No. 1.57e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 mekfmaei 8
|||||:
QY 1 MEKFMFEFG 8

RESULT 14

ID R36268 standard; Protein; 321 AA.
AC R36268;
DT 19-JAN-1994 (first entry)
DE Carnitine dehydrogenase.
KW CDHase; clinical measurement.
OS Alkaligenes sp. NO. 981 FERM BP-2570.
PN J05161492-A.
PD 29-JUN-1993.
PF 08-JUL-1991; 193471.
PR 08-JUL-1991; JP-193471.
PA (ASAH) ASAH CHEM IND CO LTD.
DR WPI; 93-239257/30.

DR N-PSDB; Q46585.
PT New purified Escherichia coli producing carnitine dehydrogenase -
PT useful for clinical measurement of carnitine
PS Disclosure; Pages 11-13; 21pp; Japanese.
CC The sequence is that of carnitine dehydrogenase (CDHase) which
CC may be produced in E. coli transformed with an expression plasmid
CC containing the CDHase gene without the addition of carnitine. CDHase
CC is useful for clinical measurement of carnitine.
SQ Sequence 321 AA;

Query Match 41.3%; Score 57; DB 8; Length 321;
Best Local Similarity 50.0%; Pred. No. 1.57e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

Db 246 mrhfmaeqfpa-lqlp 260
|:|||||:
QY 1 MEKFMFEFGQGYVQTP 16

RESULT 15

ID W72197 standard; Protein; 623 AA.
AC W72197;
DT 13-JAN-1999 (first entry)
DE HSV-2 strain SB5 Contig ID 15 ORF#32 protein.
KW HSV-2 strain SB5; immunological response induction; therapy;
KW antiviral identification; viral protein inhibitor.
OS Herpes simplex virus type 2.
PN WO9820016-A1.
PD 14-MAY-1998.
PF 31-OCT-1997; U20016.


```

periwinkle
DATE      13-Jan-1995 #sequence_revision 09-May-1997 #text_change
ACCESSIONS S35168
REFERENCE   S35168
#authors    Meijer, A.H.; Souer, E.; Verpoorte, R.; Hoge, J.H.C.
#journal     Plant Mol. Biol. (1993) 22:379-383
#title       Isolation of cytochrome P-450 cDNA clones from the higher
#cross-references MUID:93283641
#accession     S35168
#molecule_type mRNA
#residues      1-56 ##label MEI
##cross-references EMBL:X69775; NID:g395301; PID:g395302
CLASSIFICATION #superfamily unassigned cytochrome P450; cytochrome P450
homology
KEYWORDS     chromoprotein; electron transfer; heme; iron; monooxygenase;
oxidoreductase
FEATURE      2
#binding_site heme iron (Cys) (axial ligand) #status
predicted
SUMMARY      #length 56 #checksum 2986
Query Match 47.8%; Score 66; DB 2; Length 56;
Best Local Similarity 29.4%; Pred. No. 2.24e+00;
Matches 5; Conservative 9; Mismatches 3; Indels 0; Gaps 0;
Db 20 LOREKEDVAPSYVHAPF 36
:::|: : :|:|:|
QY 1 MEKFMAEFGQGYVQTPF 17

RESULT      3
ENTRY       I64032 #type complete
TITLE       hypothetical protein H11500 - Haemophilus influenzae (strain
ORGANISM    Rd KW20)
DATE        18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
ACCESSIONS I64032
REFERENCE   A64000
#authors    Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.;
Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
Fitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,
J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
#journal     Science (1995) 269:496-512
#title       Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd.
#cross-references MUID:95350630
#accession   I64032
#status      nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues    1-508 ##label TIGR
##cross-references GB:U32826; GB:L42023; NID:g1574322; PID:g1574341;
TIGR:H11500
SUMMARY      #length 508 #molecular-weight 57190 #checksum 7619
Query Match 47.8%; Score 66; DB 2; Length 508;
Best Local Similarity 43.8%; Pred. No. 2.24e+00;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
Db 411 MPKYALYESGYTQIP 426
| |::| : |::|
QY 1 MEKFMAEFGQGYVQTP 16

RESULT      4

```

```

ENTRY       T02323 #type complete
TITLE       hypothetical protein F13P17.19 - Arabidopsis thaliana
ORGANISM    #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE        05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change
ACCESSIONS T02323
REFERENCE   Z14168
#authors    Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.;
Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.;
Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter,
J.C.
#submission submitted to the EMBL Data Library, July 1998
#description Arabidopsis thaliana chromosome II BAC F13P17 genomic
sequence.
#accession   T02323
#status      preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues   1-2301 ##label ROU
##cross-references EMBL:AC004481; NID:g3337347; PID:g3337366
GENETICS
#map_position 2
#introns      155/3; 297/1; 525/2; 550/1; 609/1; 767/3; 911/1; 1139/2;
1238/3; 1412/1; 1498/1; 1528/3; 1598/3; 1732/3; 1791/2;
1822/3; 1841/3; 1884/3; 1907/3; 1945/3; 1999/3; 2033/3;
2060/2; 2107/2
#note        F13P17.19
SUMMARY      #length 2301 #molecular-weight 253398 #checksum 4265
Query Match 47.1%; Score 65; DB 2; Length 2301;
Best Local Similarity 46.7%; Pred. No. 3.25e+00;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
Db 977 NFLGRFGAGYVDTF 991
:::|:| |::|
QY 3 KFMFAEFGQGYVQTPF 17

RESULT      5
ENTRY       D70428 #type complete
TITLE       hypothetical protein aq_1477 - Aquifex aeolicus
ORGANISM    #formal_name Aquifex aeolicus
DATE        08-May-1998 #sequence_revision 08-May-1998 #text_change
ACCESSIONS D70428
REFERENCE   A70300
#authors    Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;
Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short,
J.M.; Olson, G.J.; Swanson, R.V.
#journal     Nature (1998) 392:353-358
#title       The complete genome of the hyperthermophilic bacterium
Aquifex aeolicus.
#cross-references MUID:98196666
#accession   D70428
#status      preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues    1-273 ##label AQF
##cross-references GB:AE000742; NID:g2983858; PID:g2983869; GB:AE000657
##experimental_source strain VF5
GENETICS
#gene        aq_1477
SUMMARY      #length 273 #molecular-weight 32949 #checksum 7504
Query Match 46.4%; Score 64; DB 2; Length 273;
Best Local Similarity 38.5%; Pred. No. 4.70e+00;
Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;
Db 82 QRELDERVEGYID 94
:::|:| |::|
QY 2 EKFMFAEFGQGYVQ 14

```

```
RESULT 6
ENTRY TVUTG4 #type complete
TITLE phosphoglycerate kinase (EC 2.7.2.3), cytosolic (allele 4) -
ALTERNATE_NAMES Trypanosoma brucei
ORGANISM #formal_name Trypanosoma brucei
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
ACCESSIONS S02234
REFERENCE S00748
#authors le Blancq, S.M.; Swinkels, B.W.; Gibson, W.C.; Borst, P.
#journal J. Mol. Biol. (1988) 200:439-447
#title Evidence for gene conversion between the phosphoglycerate
#cross-references MUID:88286728
#accession S02234
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-420 #label LEB
#note the nucleotide sequence was submitted to the EMBL Data
LIBRARY, September 1987
CLASSIFICATION #superfamily phosphoglycerate kinase
KEYWORDS ATP; gluconeogenesis; glycolysis; phosphotransferase
FEATURE 223,345
SUMMARY #length 420 #molecular-weight 45087 #checksum 8922
Query Match 46.4%; Score 64; DB 1; Length 420;
Best Local Similarity 35.3%; Pred. No. 4.70e+00;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 151 MAKILASYGDVYISDAF 167
| | : | : | : | : |
QY 1 MEKFMFEFGGYVQTFP 17

RESULT 7
ENTRY KIUTGG #type complete
TITLE phosphoglycerate kinase (EC 2.7.2.3), glycosomal (allele 2) -
ALTERNATE_NAMES Trypanosoma brucei
ORGANISM #formal_name Trypanosoma brucei
DATE 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change
ACCESSIONS B25119
REFERENCE A25119
#authors Osinga, K.A.; Swinkels, B.W.; Gibson, W.C.; Borst, P.;
#journal Veeneman, G.H.; Van Boom, J.H.; Michels, P.A.M.; Opperdoes,
#title F.R.
#cross-references MUID:86136022
#accession B25119
#molecule_type DNA
#residues 1-440 #label OSI
#cross-references GB:X03370; NID:g10496; PID:g10498
COMMENT This is one of the three isozymes of T. brucei phosphoglycerate
kinase; it is transported to glycosomes, which belong to a group
of subcellular organelles called microbodies.
CLASSIFICATION #superfamily phosphoglycerate kinase
KEYWORDS ATP; gluconeogenesis; glycolysis; glycosome;
phosphotransferase
FEATURE 223,345
SUMMARY #binding_site ATP (Lys, Glu) #status predicted
#length 440 #molecular-weight 47118 #checksum 1365
Query Match 46.4%; Score 64; DB 1; Length 440;
Best Local Similarity 35.3%; Pred. No. 4.70e+00;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
```

```
Db 151 MAKILASYGDVYISDAF 167
| | : | : | : | : |
QY 1 MEKFMFEFGGYVQTFP 17

RESULT 8
ENTRY TVUTGB #type complete
TITLE phosphoglycerate kinase (EC 2.7.2.3), glycosomal (allele 4) -
ALTERNATE_NAMES Trypanosoma brucei
ORGANISM #formal_name Trypanosoma brucei
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
ACCESSIONS S02235
REFERENCE S00748
#authors le Blancq, S.M.; Swinkels, B.W.; Gibson, W.C.; Borst, P.
#journal J. Mol. Biol. (1988) 200:439-447
#title Evidence for gene conversion between the phosphoglycerate
#cross-references MUID:88286728
#accession S02235
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-440 #label LEB
#note the nucleotide sequence was submitted to the EMBL Data
LIBRARY, September 1987
CLASSIFICATION #superfamily phosphoglycerate kinase
KEYWORDS ATP; glycolysis; glycosome; phosphotransferase
FEATURE 223,345
SUMMARY #binding_site ATP (Lys, Glu) #status predicted
#length 440 #molecular-weight 47176 #checksum 1311
Query Match 46.4%; Score 64; DB 1; Length 440;
Best Local Similarity 35.3%; Pred. No. 4.70e+00;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 151 MAKILASYGDVYISDAF 167
| | : | : | : | : |
QY 1 MEKFMFEFGGYVQTFP 17

RESULT 9
ENTRY TVUT2B #type complete
TITLE phosphoglycerate kinase (EC 2.7.2.3) A (allele 2) -
ALTERNATE_NAMES Trypanosoma brucei
ORGANISM #formal_name Trypanosoma brucei
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
ACCESSIONS S00748
REFERENCE S00748
#authors le Blancq, S.M.; Swinkels, B.W.; Gibson, W.C.; Borst, P.
#journal J. Mol. Biol. (1988) 200:439-447
#title Evidence for gene conversion between the phosphoglycerate
#cross-references MUID:88286728
#accession S00748
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-505 #label LEB
#cross-references EMBL:X05889; NID:g10488; PID:g10489
#note the nucleotide sequence was submitted to the EMBL Data
LIBRARY, September 1987
CLASSIFICATION #superfamily phosphoglycerate kinase
KEYWORDS ATP; glycolysis; phosphotransferase
FEATURE 312,434
SUMMARY #binding_site ATP (Lys, Glu) #status predicted
#length 505 #molecular-weight 55375 #checksum 4355
Query Match 46.4%; Score 64; DB 1; Length 505;
Best Local Similarity 35.3%; Pred. No. 4.70e+00;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
```

Db 240 MAKILASYGDVYISDAF 256
| : : | : : | : : |
QY 1 MEKFMAEFGGYVQTFF 17

RESULT 10
ENTRY TVUT4B #type complete
TITLE phosphoglycerate kinase (EC 2.7.2.3) A (allele 4) -
ORGANISM Trypanosoma brucei
DATE #formal_name Trypanosoma brucei
05-Sep-1997 #sequence_revision 31-Dec-1991 #text_change
S02233
ACCESSIONS S00748
REFERENCE le Blancq, S.M.; Swinkels, B.W.; Gibson, W.C.; Borst, P.
#authors J. Mol. Biol. (1988) 200:439-447
#journal
#title Evidence for gene conversion between the phosphoglycerate
kinase genes of *Trypanosoma brucei*.
#cross-references MUID:88286728
#accession S02233
#molecule_type DNA
#status nucleic acid sequence not shown; translation not shown
#residues 1-508 #label LEB
#note the nucleotide sequence was submitted to the EMBL Data
Library, September 1987
CLASSIFICATION #superfamily phosphoglycerate kinase
KEYWORDS ATP; glycolysis; phosphotransferase
FEATURE
312.434 #binding_site ATP (Lys, Glu) #status predicted
SUMMARY #length 508 #molecular_weight 55668 #checksum 5682
Query Match 46.4%; Score 64; DB 1; Length 508;
Best Local Similarity 35.3%; Pred. No. 4.70e+00;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 240 MAKILASYGDVYISDAF 256
| : : | : : | : : |
QY 1 MEKFMAEFGGYVQTFF 17

RESULT 11
ENTRY A45593 #type complete
TITLE phosphoglycerate kinase (EC 2.7.2.3), glycosomal -
ORGANISM Trypanosoma brucei
DATE #formal_name Trypanosoma brucei
22-Jan-1999 #sequence_revision 02-Jun-1994 #text_change
A45593
ACCESSIONS Alexander, K.; Parsons, M.
REFERENCE Mol. Biochem. Parasitol. (1991) 46:1-10
#authors A phosphoglycerate kinase-like molecule localized to
#journal glycosomal microbodies: evidence that the topogenic signal
#title is not at the C-terminus.
#cross-references MUID:91304507
#accession A45593
#molecule_type DNA
#residues 1-509 #label ALE
#cross-references GB:M37784; NID:G295366; PID:G295367
#experimental_source EATRO 164
#note sequence extracted from NCBI backbone (NCBIN:42855,
NCBIP:42870)
CLASSIFICATION #superfamily phosphoglycerate kinase
KEYWORDS ATP; glycosome; phosphotransferase
SUMMARY #length 509 #molecular_weight 55747 #checksum 9972
Query Match 46.4%; Score 64; DB 1; Length 509;
Best Local Similarity 35.3%; Pred. No. 4.70e+00;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 240 MAKILASYGDVYISDAF 256
| : : | : : | : : |

QY 1 MEKFMAEFGGYVQTFF 17
| : : | : : | : : |

RESULT 12
ENTRY JS0443 #type complete
TITLE alanine racemase (EC 5.1.1.1) - *Bacillus subtilis*
ORGANISM #formal_name *Bacillus subtilis*
DATE 07-Sep-1990 #sequence_revision 23-May-1997 #text_change
05-Dec-1998
ACCESSIONS JS0443; G69612
REFERENCE JS0443
#authors Ferrari, E.; Henner, D.J.; Yang, M.Y.
#journal Bio/technology (1985) 3:1003-1007
#title Isolation of an alanine racemase gene from *Bacillus subtilis*
and its use for plasmid maintenance in *B. subtilis*.
#accession JS0443
#molecule_type DNA
#residues 1-389 #label FER
#cross-references EMBL:M16207; NID:G142821; PID:G142822
#accession A69580
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.;
Guseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Hoisappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karanata, D.;
Kashara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Maueel, C.; Medique, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rev, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession G69612
#status nucleic acid sequence not shown; translation not shown
#residues 1-39, 'V', 41-65, 'V', 67-389 #label KUN
#cross-references GB:299106; GB:AL009126; NID:G2632653; PID:ell82430;
PID:G2632764
#experimental_source strain 168
GENETICS
#gene dal
#gene catalyzes the racemic interconversion of L-alanine and
FUNCTION D-alanine
#description cell wall synthesis
#pathway

```
CLASSIFICATION #superfamily alanine racemase
KEYWORDS cell wall synthesis; isomerase; phosphoprotein; pyridoxal
          phosphate
FEATURE
41 #binding_site pyridoxal phosphate (Lys) (covalent)
SUMMARY #length 389 #molecular-weight 43327 #checksum 9391

Query Match 45.7%; Score 63; DB 1; Length 389;
Best Local Similarity 50.0%; Pred. No. 6.76e+00;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 313 MDQWVVDQY 324
| : ||| :| :| :|
Qy 1 MEKFAEFGQGYVTPPF 12

RESULT 13
ENTRY TVCRGG #type complete
TITLE phosphoglycerate kinase (EC 2.7.2.3), cytosolic - Crithidia
ALTERNATE_NAMES fasciculata
ORGANISM #phosphoglycerate kinase B
#formal_name Crithidia fasciculata
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
05-Sep-1997
ACCESSIONS S00486
REFERENCE Swinkels, B.W.; Evers, R.; Borst, P.
EMBO J. (1988) 7:1159-1165
#journal The topogenic signal of the glycosomal (microbody)
#title phosphoglycerate kinase of Crithidia fasciculata resides in
a carboxy-terminal extension.
#cross-references MUID:88296420
#accession S00486
#molecule_type DNA
#residues 1-417 #label SWI
#cross-references EMBL:X07458; NID:g6982; PID:g6983
#note the authors translated the codon GAC for residue 173 as
293 as Ala

CLASSIFICATION #superfamily phosphoglycerate kinase
KEYWORDS ATP; gluconeogenesis; glycolysis; phosphotransferase
FEATURE
220,342 #binding_site ATP (Lys, Glu) #status predicted
SUMMARY #length 417 #molecular-weight 44602 #checksum 3677

Query Match 45.7%; Score 63; DB 1; Length 417;
Best Local Similarity 41.2%; Pred. No. 6.76e+00;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Db 148 MAKVLAAYGDVYVSDAF 164
| :| :| :| :| :|
Qy 1 MEKFAEFGQGYVTPPF 17

RESULT 14
ENTRY TVCRGG #type complete
TITLE phosphoglycerate kinase (EC 2.7.2.3), glycosomal - Crithidia
ALTERNATE_NAMES fasciculata
ORGANISM #phosphoglycerate kinase C
#formal_name Crithidia fasciculata
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
05-Sep-1997
ACCESSIONS S00487
REFERENCE Swinkels, B.W.; Evers, R.; Borst, P.
EMBO J. (1988) 7:1159-1165
#journal The topogenic signal of the glycosomal (microbody)
#title phosphoglycerate kinase of Crithidia fasciculata resides in
a carboxy-terminal extension.
#cross-references MUID:88296420
#accession S00487
#molecule_type DNA
```

```
#residues 1-455 #label SWI
#cross-references EMBL:X07459; NID:g6984; PID:g6985
#note the authors translated the codon GAC for residue 173 as
thr, CAA for residue 292 as Lys, and GGC for residue
293 as Ala
CLASSIFICATION #superfamily phosphoglycerate kinase
KEYWORDS ATP; glycolysis; glycosome; phosphotransferase
FEATURE
220,342 #binding_site ATP (Lys, Glu) #status predicted
SUMMARY #length 455 #molecular-weight 47843 #checksum 8443

Query Match 45.7%; Score 63; DB 1; Length 455;
Best Local Similarity 41.2%; Pred. No. 6.76e+00;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Db 148 MAKVLAAYGDVYVSDAF 164
| :| :| :| :| :|
Qy 1 MEKFAEFGQGYVTPPF 17

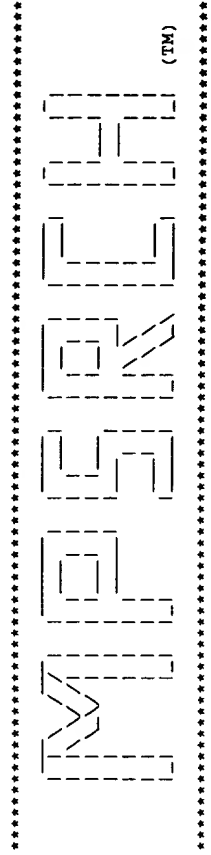
RESULT 15
ENTRY E71569 #type complete
TITLE probable acyltransferase - Chlamydia trachomatis (serotype D,
strain UM3/Cx)
ORGANISM #formal_name Chlamydia trachomatis
DATE 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change
21-Nov-1998
ACCESSIONS E71569
REFERENCE A71570
#authors Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe,
R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov,
R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
#journal Science (1998) 282:754-759
#title Genome sequence of an obligate intracellular pathogen of
humans: Chlamydia trachomatis.
#cross-references MUID:99000809
#accession E71569
#status preliminary
#molecule_type DNA
#residues 1-455 #label ARN
#cross-references GB:AE001275; GB:AE001273; NID:g3328388; PID:g3328398
#experimental_source serotype D, strain UM-3/Cx

GENETICS
#gene htrB
SUMMARY #length 455 #molecular-weight 52058 #checksum 3772

Query Match 45.7%; Score 63; DB 2; Length 455;
Best Local Similarity 29.4%; Pred. No. 6.76e+00;
Matches 5; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

Db 343 LQRFLEFGEFYADASL 359
| :| :| :| :| :| :|
Qy 1 MEKFAEFGQGYVTPPF 17

Search completed: Wed Sep 1 16:08:44 1999
Job time : 12 secs.
```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:07:26 1999; Maspar time 2.71 Seconds
Tabular output not generated. 177.194 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (1-17) from PCTUS9913024.pap (3 of 12)
Perfect Score: 138
Sequence: 1 MEKFMFEGQGVQTFP 17

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 29.668; Variance 42.415; scale 0.699

* Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES							
Result	Query	Match	Length	DB	ID	Description	Pred. No.
1	66	47.8	508	1	YF00_HAEIN	HYPOTHETICAL PROTEIN H	7.70e-01
2	66	47.8	524	1	CP72_CATRO	CYTOTOXIC P450 72A1 (7.70e-01
3	64	46.4	420	1	PGKE_TRYBB	PHOSPHOGLYCERATE KINAS	1.73e+00
4	64	46.4	440	1	PGKC_TRYBB	PHOSPHOGLYCERATE KINAS	1.73e+00
5	64	46.4	505	1	PGKA_TRYBB	PHOSPHOGLYCERATE KINAS	1.73e+00
6	64	46.4	508	1	PGKD_TRYBB	PHOSPHOGLYCERATE KINAS	1.73e+00
7	63	45.7	389	1	ALR_BACSU	ALANINE RACEMASE (EC 5	2.59e+00
8	63	45.7	417	1	PGKB_CRIFA	PHOSPHOGLYCERATE KINAS	2.59e+00
9	63	45.7	455	1	PGKC_CRIFA	PHOSPHOGLYCERATE KINAS	2.59e+00
10	63	45.7	603	1	VEL_HPVO8	REPLICATION PROTEIN E1	2.59e+00
11	62	44.9	417	1	PGKB_LEIMA	PHOSPHOGLYCERATE KINAS	3.84e+00
12	62	44.9	421	1	PGKB_TRYBB	PHOSPHOGLYCERATE KINAS	3.84e+00
13	62	44.9	479	1	PGKC_LEIMA	PHOSPHOGLYCERATE KINAS	3.84e+00
14	62	44.9	604	1	VEL_HPVS6	REPLICATION PROTEIN E1	3.84e+00
15	62	44.9	858	1	UBP5_HUMAN	UBIQUITIN CARBOXYL-TER	3.84e+00
16	62	44.9	858	1	UBP5_MOUSE	UBIQUITIN CARBOXYL-TER	3.84e+00
17	61	44.2	476	1	VTDB_RABIT	VITAMIN D-BINDING PROT	5.67e+00
18	61	44.2	605	1	VEL_HPVO9	REPLICATION PROTEIN E1	5.67e+00
19	60	43.5	734	1	SULX_YEAST	POTASSIUM SULFATE TRANS	8.34e+00
20	60	43.5	1010	1	Y661_HAEIN	PROBABLE TONB-DEPENDEN	8.34e+00
21	59	42.8	163	1	YMJ4_YEAST	HYPOTHETICAL 18.4 KD P	1.22e+01
22	59	42.8	262	1	YGE6_YEAST	HYPOTHETICAL 29.4 KD P	1.22e+01
23	59	42.8	367	1	GLDA_ECOLI	GLYCEROL DEHYDROGENASE	1.22e+01

RESULT	ID	YF00_HAEIN	STANDARD;	PRT;	508 AA.
AC	P44224;				
DT	01-NOV-1995	(REL. 32, CREATED)			
DT	01-NOV-1995	(REL. 32, LAST SEQUENCE UPDATE)			
DT	01-NOV-1995	(REL. 32, LAST ANNOTATION UPDATE)			
DE	HYPOTHETICAL PROTEIN H11500.				
GN	H11500.				
OS	HAEMOPHILUS INFLUENZAE.				
OC	BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;				
OC	HAEMOPHILUS.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-RD / KW20;				
RX	MEDLINE: 95350630.				
RA	FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,				
RA	KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,				
RA	MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,				
RA	SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,				
RA	WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., COTTON M.D.,				
RA	UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,				
RA	FINE L.D., FRITZMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,				
RA	GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,				
RA	VENTER J.C.;				
RT	"Whole-genome random sequencing and assembly of Haemophilus				
RT	Influenzae Rd.;"				
RL	SCIENCE 269:496-512(1995).				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
DR	EMBL; U32826; G1574341; "				
DR	TIGR; H11500; "				
KW	HYPOTHETICAL PROTEIN.				
SQ	SEQUENCE 508 AA; 57190 MW; 98DDC6AA CRC32;				

ALIGNMENTS

Query Match	Length	DB	ID	Description	Pred. No.
1	66	47.8	508	1	YF00_HAEIN
2	66	47.8	524	1	CP72_CATRO
3	64	46.4	420	1	PGKE_TRYBB
4	64	46.4	440	1	PGKC_TRYBB
5	64	46.4	505	1	PGKA_TRYBB
6	64	46.4	508	1	PGKD_TRYBB
7	63	45.7	389	1	ALR_BACSU
8	63	45.7	417	1	PGKB_CRIFA
9	63	45.7	455	1	PGKC_CRIFA
10	63	45.7	603	1	VEL_HPVO8
11	62	44.9	417	1	PGKB_LEIMA
12	62	44.9	421	1	PGKB_TRYBB
13	62	44.9	479	1	PGKC_LEIMA
14	62	44.9	604	1	VEL_HPVS6
15	62	44.9	858	1	UBP5_HUMAN
16	62	44.9	858	1	UBP5_MOUSE
17	61	44.2	476	1	VTDB_RABIT
18	61	44.2	605	1	VEL_HPVO9
19	60	43.5	734	1	SULX_YEAST
20	60	43.5	1010	1	Y661_HAEIN
21	59	42.8	163	1	YMJ4_YEAST
22	59	42.8	262	1	YGE6_YEAST
23	59	42.8	367	1	GLDA_ECOLI

Query Match 47.8%; Score 66; DB 1; Length 508;
Best Local Similarity 43.8%; Pred. No. 7.70e-01;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
Db 411 MPKAYKALYESGYTQIP 426
| : : : | : : : |

Qy 1 MEKFAEFGQGVOTPF 16

```
RESULT 2
ID CP72_CATRO STANDARD; PRT; 524 AA.
AC Q05047;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450 72A1 (EC 1.14.14.1) (CYPLXXII) (PROBABLE GERANIOL-10-
DE HYDROXYLASE) (GEL0H).
GN CYP72A1 OR CYP72 OR P450C83.
OS CATHARANTHUS ROSEUS (ROSY PERIWINKLE) (MADAGASCAR PERIWINKLE).
OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; EMERYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTES; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; GENTIANACEAE; GENTIANALES; APOCYNACEAE; CATHARANTHUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LINE CP3A;
RA VETTER H.-P., MANGOLD U., SCHROEDER G., MARNER F.-J.,
RA WERCK-REICHART D., SCHROEDER J.;
RT "Molecular analysis and heterologous expression of an inducible
RT cytochrome P-450 protein from periwinkle (Catharanthus roseus L.).";
RL PLANT PHYSIOL. 100:998-1007(1992).
RN [2]
RP SEQUENCE OF 469-524 FROM N.A.
RX MEDLINE; 93283641.
RA MEIJER A.H., SOUVER E., VERPOORTE R., HOGE J.H.C.;
RT "Isolation of cytochrome P-450 cDNA clones from the higher plant
RT Catharanthus roseus by a PCR strategy.";
RL PLANT MOL. BIOL. 22:379-383(1993).
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. MAY BE A GERANIOL-10-HYDROXYLASE.
CC -!- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +
CC OXIDIZED FLAVOPROTEIN + H(2)O.
CC -!- PATHWAY: INDOLE ALKALOID SYNTHESIS.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L10081; G167484; -
CC EMBL; X69775; G395302; -
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC PFAM; PF00067; P450; 1.
CC KX OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; TRANSMEMBRANE;
CC HEME.
CC TRANSMEM 12 32 POTENTIAL.
CC BINDING 470 470 HEME (BY SIMILARITY).
CC VARIANT 190 190 I -> L.
CC VARIANT 194 194 Q -> E.
CC VARIANT 223 223 E -> D.
CC VARIANT 312 312 K -> R.
CC VARIANT 318 318 S -> T.
CC VARIANT 403 403 V -> I.
CC VARIANT 405 405 K -> E.
CC VARIANT 411 411 S -> P.
CC SEQUENCE 524 AA; 60557 MW; FC0286BC CRC32;
Query Match 47.8%; Score 66; DB 1; Length 524;
Best Local Similarity 29.4%; Pred. No. 7,70e-01;
Matches 5; Conservative 9; Mismatches 3; Indels 0; Gaps 0;
```

Db 488 LQRFKFDVAPSYVHAPF 504

Qy 1 MEKFAEFGQGVOTPF 17

```
RESULT 3
ID PGKC_TRYBB STANDARD; PRT; 420 AA.
AC P08893;
DT 01-NOV-1988 (REL. 09, CREATED)
DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE PHOSPHOGLYCERATE KINASE, CYTOSOLIC (B) (EC 2.7.2.3) (ALLELE 4).
OS TRYPANOSOMA BRUCEI BRUCEI.
OC EUKARYOTA; EULENZOEA; KINETOPLASTIDA; TRYPANOSOMATIDAE; TRYPANOSOMA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88286728.
RA LE BLANCQ S.M., SWINKELS B.W., GIBSON W.C., BORST P.;
RT "Evidence for gene conversion between the phosphoglycerate kinase
RT genes of Trypanosoma brucei.";
RL J. MOL. BIOL. 200:439-447(1988).
CC -!- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE -> ADP +
CC 3-PHOSPHO-D-GLYCEROYL PHOSPHATE.
CC -!- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -!- SUBUNIT: MONOMER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- IN T. BRUCEI, THREE GENES CODE FOR PHOSPHOGLYCERATE KINASE
CC ISOZYMES, WHICH ARE TRANSPORTED TO DIFFERENT CELL COMPARTMENTS.
CC -!- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X05890; G10494; -
CC PRS; S02334; TVUTG4.
CC PROSITE; PS00111; GLYCERATE_KINASE; 1.
CC PFAM; PF00162; PGK; 1.
CC DR HSPP; P07378; 13PK.
CC KW TRANSFERASE; KINASE; GLYCOLYSIS; MULTIGENE FAMILY.
CC SEQUENCE 420 AA; 45087 MW; DDFAJ3B80 CRC32;
Query Match 46.4%; Score 64; DB 1; Length 420;
Best Local Similarity 35.3%; Pred. No. 1.73e+00;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
Db 151 MAKILASGVYVISDAF 167
Qy 1 MEKFAEFGQGVOTPF 17

RESULT 4
ID PGKC_TRYBB STANDARD; PRT; 440 AA.
AC P07378;
DT 01-APR-1988 (REL. 07, CREATED)
DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PHOSPHOGLYCERATE KINASE, GLYCOSOMAL (C) (EC 2.7.2.3).
OS TRYPANOSOMA BRUCEI BRUCEI.
OC EUKARYOTA; EULENZOEA; KINETOPLASTIDA; TRYPANOSOMATIDAE; TRYPANOSOMA.
RN [1]
RP SEQUENCE FROM N.A. (ALLELES 2 AND 4).
RX MEDLINE; 88286728.
RA LE BLANCQ S.M., SWINKELS B.W., GIBSON W.C., BORST P.;
RT "Evidence for gene conversion between the phosphoglycerate kinase
RT genes of Trypanosoma brucei.";
RL J. MOL. BIOL. 200:439-447(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86136022.
RA OSINGA K.A., SWINKELS B.W., GIBSON W.C., BORST P., VEENEMAN G.H.,
RA VAN BOOM J.H., MICHELIS P.A.M., OPPERDOES F.R.;
RT "Topogenesis of microbody enzymes: a sequence comparison of the genes
```

```

J. MOL. BIOL. 200:439-447(1988).
CC -1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE = ADP +
CC 3-PHOSPHO-D-GLYCEROYL PHOSPHATE.
CC -1- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: NOT KNOWN.
CC -1- IN T. BRUCEI, THREE GENES CODE FOR PHOSPHOGLYCERATE KINASE
CC ISOZYMES, WHICH ARE TRANSPORTED TO DIFFERENT CELL COMPARTMENTS.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X05889; GI0489; -.
DR DR PIR; S00748; TVUT28.
DR DR PROSITE; PS00111; GLYCERATE_KINASE; 1.
DR DR PFAM; PF00162; PGK; 1.
DR DR HSP; P07378; 13PK.
DR DR TRANSFERASE; KINASE; GLYCOLYSIS; MULTIGENE FAMILY.
SQ SEQUENCE 505 AA; 55376 MW; 9D2ACC09 CRC32;
-----
Query Match 46.48; Score 64; DB 1; Length 505;
Best Local Similarity 35.38; Pred. No. 1.73e+00; Indels 0; Gaps 0;
Matches 6; Conservative 6; Mismatches 5;
-----
Db 240 MAKILASYGDVYISDAF 256
| | : | : | : | : |
QY 1 MEKFMFEGGQYVQTF 17
-----
RESULT 6
ID ID PGKD_TRYBB STANDARD; PRT; 508 AA.
AC P08892;
DT 01-NOV-1988 (REL. 09, CREATED)
DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)
DE PHOSPHOGLYCERATE KINASE, A (EC 2.7.2.3) (ALLELE 4).
OS TRYPA NOSOMA BRUCEI BRUCEI.
TX EUKARYOTA; EUKLEOZOIA; KINETOPLASTIDA; TRYPA NOSOMATIDAE; TRYPA NOSOMA.
[1]
SEQUENCE FROM N.A.
RX MEDLINE; 88286728.
RA LE BLANCQ S.M., SWINKELS B.W., GIBSON W.C., BORST P.;
RT "Evidence for gene conversion between the phosphoglycerate kinase
RT genes of Trypanosoma brucei."
J. MOL. BIOL. 200:439-447(1988).
CC -1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE = ADP +
CC 3-PHOSPHO-D-GLYCEROYL PHOSPHATE.
CC -1- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: NOT KNOWN.
CC -1- IN T. BRUCEI, THREE GENES CODE FOR PHOSPHOGLYCERATE KINASE
CC ISOZYMES, WHICH ARE TRANSPORTED TO DIFFERENT CELL COMPARTMENTS.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X05890; GI0493; -.
DR DR PIR; S02233; TVUT48.
DR DR PROSITE; PS00111; GLYCERATE_KINASE; 1.
DR DR PFAM; PF00162; PGK; 1.
DR DR HSP; P07378; 13PK.

```



```
CC WHICH IS NOT PRESENT IN THE CYTOSOLIC ISOENZYME. THIS DOMAIN MOST
CC LIKELY SERVES AS TOPOGENIC SIGNAL TO DIRECT THE GLYCOSOMAL PKG TO
CC THE GLYCOSOME.
CC -!- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X07459; G6985; -.
DR PIR; S00487; TVCRGG.
DR DR PROSITE; PS00111; PGLYCERATE_KINASE; 1.
DR PFAM; PF00162; PGK; 1.
DR HSP; P07378; 13PK.
KW TRANSFERASE; KINASE; GLYCOLYSIS; MULTIGENE FAMILY; GLYCOSOME.
FT DOMAIN 417 455 TOPOGENIC SIGNAL (POTENTIAL).
FT DOMAIN 422 427 POLY-A.
SQ SEQUENCE 455 AA; 47843 MW; C5062442 CRC32;
-----
Query Match 45.7%; Score 63; DB 1; Length 455;
Best Local Similarity 41.28; Pred. No. 2.59e+00;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
Db 148 MAKVLAAYGDVVVSDF 164
QY 1 MEKFMAEFGGYQTFF 17
| | : : : : | : |
RESULT 10
ID VEL_HPV08 STANDARD; PRT; 603 AA.
AC P06420;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DT REPLICATION PROTEIN EI.
DE EN.
OS HUMAN PAPILLOMAVIRUS TYPE 8.
GN VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PAPOVAVIRIDAE; PAPILLOMAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86200410.
RA FUCHS P.G., IFTNER T., WENINGER J., PRISTER H.;
RT "Epidermodysplasia verruciformis-associated human papillomavirus 8:
RT genomic sequence and comparative analysis.";
RL J. VIROL. 58:626-634(1986).
CC -!- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF
CC VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2
CC PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH
CC CONTAINS BINDING SITES FOR BOTH PROTEINS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M12737; -: NOT_ANNOTATED_CDS.
DR PIR; A03656; WLWL8.
DR PFAM; PF00519; EI; 1.
DR PFAM; PF00524; EI_LN; 1.
KW EARLY PROTEIN; DNA REPLICATION; HELICASE; ATP-BINDING; DNA-BINDING;
KW NUCLEAR PROTEIN.
FT NP_BIND 431 438 ATP (POTENTIAL).
SQ SEQUENCE 603 AA; 68821 MW; E7BCC22E CRC32;
-----
Query Match 45.7%; Score 63; DB 1; Length 603;
```


OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE; 96105388.
RA FALQUET L., PAQUET N., FRUTIGER S., HUGHES G.J., HOANG-VAN K.,
RA JATON J.-C.;
RT "cDNA cloning of a human 100 kDa de-ubiquitinating enzyme: the 100
RT kDa human de-ubiquitinase belongs to the ubiquitin C-terminal
RT hydrolase family 2 (UCH2).";
RL FEBS LETT. 376:233-237(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96303695.
RA ANSARI-LARI M.A., MUZNY D.M., LU J., LU F., LILLEY C.E., SPANOS S.,
RA MALLEY T., GIBBS R.A.;
RT "A gene-rich cluster between the CD4 and triosephosphate isomerase
RT genes at human chromosome 12p13.";
RN [3]
RP SEQUENCE FROM N.A.
RX TASHAYEV V.L., O'CONNOR L.B., LARSEN C.N., KASPEREK E., PICKART C.M.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP CHARACTERIZATION.
RX MEDLINE; 95154450.
RA FALQUET L., PAQUET N., FRUTIGER S., HUGHES G.J., HOANG-VAN K.,
RA JATON J.-C.;
RT "A human de-ubiquitinating enzyme with both isopeptidase and
RT peptidase activities in vitro.";
RL FEBS LETT. 359:73-77(1995).
CC -!- FUNCTION: CLEAVES LINEAR AND BRANCHED MULTIUBIQUITIN POLYMERS
CC -!- WITH A MARKED PREFERENCE FOR BRANCHED POLYMERS.
CC -!- CATALYTIC ACTIVITY: UBIQUITIN C-TERMINAL THIOLESTER + H(2)O -
CC UBIQUITIN + A THIOL.
CC -!- COFACTOR: ZINC; REQUIRED FOR CATALYTIC ACTIVITY.
CC -!- PTM: THE N-TERMINAL IS BLOCKED.
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS ARE PRODUCED BY ALTERNATIVE
CC SPLICING.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19; ALSO KNOWN AS
CC FAMILY 2 OF UBIQUITIN CARBOXYL-TERMINAL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X91349; E208113; -
DR EMBL; U47927; G1208744; -
DR EMBL; U47924; G1203896; -
DR EMBL; U47924; G1203895; -
DR EMBL; U35116; G1008542; -
DR MIM; 601447; -
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PFAM; PF00442; UCH-1; 1.
DR PFAM; PF00443; UCH-2; 1.
DR PFAM; PF00627; UBA; 2.
KW UBIQUITIN CONJUGATION; HYDROLASE; THIOL PROTEASE; MULTIGENE FAMILY;
KW ALTERNATIVE SPLICING; ZINC.
FT ACT_SITE 335 335 BY SIMILARITY.
FT ACT_SITE 809 809 BY SIMILARITY.
FT ACT_SITE 818 818 BY SIMILARITY.
FT VARSPLIC 629 652 GSLGFVGNEDSFCSPPSPPTS -> A (IN SHORT
FT FORM).
FT EL -> DV (IN REF. 1).
FT CONFLICT 3 4
FT CONFLICT 45 45 I -> V (IN REF. 1).
FT CONFLICT 468 468 K -> R (IN REF. 3).
FT CONFLICT 681 681 G -> D (IN REF. 3).
SQ SEQUENCE 858 AA; 95786 MW; 1F0B83C9 CRC32;

Query Match 44.9%; Score 62; DB 1; Length 858;
Best Local Similarity 41.2%; Pred. No. 3.84e+00;
Matches 7; Conservative 4; Mismatches 5; Indels 1;
Gaps 1;
Db 47 MNTFLG-FGKQYVERHF 62
Qy 1 MERFMAEFGGQYVOTPF 17
|:|::|||
Search completed: Wed Sep 1 16:07:37 1999
Job time : 11 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```
MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on:      Wed Sep 1 16:07:53 1999;  MasPar time 5.37 Seconds
Tabular output not generated.  172.884 Million cell updates/sec
```

```
>PCT-US99-13024-2
Title:
Description: (1-17) from PCTUS9913024.pep (3 of 12)
Perfect Score: 138
Sequence: 1 MEKFMAEFGGIVOTPF 17
```

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl19

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 28.183; Variance 44.286; scale 0.636

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	77	55.8	860	1	860AA LONG HYPOTHETICA	2.01e-02
2	69	50.0	364	3	MANNOSE-1-PHOSPHATE GU	5.40e-01
3	68	49.3	178	3	HYPOTHETICAL 19.4 KD P	8.03e-01
4	66	47.8	516	10	CYTOCHROME P450 (EC 1.	1.76e+00
5	65	47.1	2301	10	F13p17.19 PROTEIN.	2.59e+00
6	64	46.4	273	2	HYPOTHETICAL 32.9 KD P	3.80e+00
7	64	46.4	420	8	PHOSPHOGLYCERATE KINAS	3.80e+00
8	64	46.4	440	8	PHOSPHOGLYCERATE KINAS	3.80e+00
9	64	46.4	509	8	PHOSPHOGLYCERATE KINAS	3.80e+00
10	64	46.4	524	10	PHOSPHOGLYCERATE KINAS	3.80e+00
11	63	45.7	355	5	CYTOCHROME P450 (EC 1.	3.80e+00
12	63	45.7	455	2	SIMILAR TO FAMILY 1 OF	5.56e+00
13	63	45.7	645	2	ACYLTRANSFERASE	5.56e+00
14	63	45.7	1215	2	PPE-FAMILY PROTEIN.	5.56e+00
15	62	44.9	191	1	DNA FOR SEROTYPE B CAP	5.56e+00
16	61	44.2	605	14	FUCULOSE-1-PHOSPHATE A	8.10e+00
17	60	43.5	205	1	E6, E7, E1, E2, E4, L2	1.17e+01
18	60	43.5	499	2	RIBONUCLEASE HII (RNHB	1.70e+01
19	60	43.5	520	2	ALGI.	1.70e+01
20	60	43.5	558	1	558AA LONG HYPOTHETICA	1.70e+01

21	60	43.5	999	2	087296	HEMOGLOBIN BINDING PRO	1.70e+01
22	60	43.5	1006	3	Q07324	CHROMOSOME PERMEASE	1.70e+01
23	60	43.5	1394	3	Q08409	ATP DEPENDENT XV READING	1.70e+01
24	59	42.8	357	5	Q20173	SIMILAR TO HUMAN PROTE	2.44e+01
25	59	42.8	448	2	Q31209	OPR23-LIKE PORIN PRECU	2.44e+01
26	59	42.8	485	2	Q59277	BETA-1,4-XYLANASE PREC	2.44e+01
27	59	42.8	639	9	Q21970	DARA PROTEIN.	2.44e+01
28	59	42.8	661	14	Q56948	ES, E7, E1, E2, E4, L2	2.44e+01
29	59	42.8	936	2	Q25141	CHROMOSOME C BIOGENESI	2.44e+01
30	59	42.8	968	3	Q87199	KINESIN MOTOR PROTEIN.	2.44e+01
31	59	42.8	1007	2	Q67026	HYPOHETICAL I18.2 KD	2.44e+01
32	59	42.8	1465	2	Q51734	HYPOTHETICAL 168.8 FRG	2.44e+01
33	58	42.0	75	2	Q54589	M TYPE 58 (EMML) (FRAG	3.50e+01
34	58	42.0	338	14	Q11416	52K PROTEIN.	3.50e+01
35	58	42.0	391	2	Q34929	YKFE PROTEIN.	3.50e+01
36	58	42.0	728	4	Q13779	APOLIPOPROTEIN B-48 (F	3.50e+01
37	58	42.0	953	5	Q19753	F23B12.7 PROTEIN.	3.50e+01
38	58	42.0	1570	3	Q14168	PROBABLE PROTEIN TRANS	3.50e+01
39	58	42.0	1783	11	Q70395	MYOSIN XV (UNCONVENTIO	3.50e+01
40	58	42.0	2219	5	Q23388	ZK1067.2 PROTEIN.	3.50e+01
41	58	42.0	3262	4	Q13788	APOLIPOPROTEIN B-100 (3.50e+01
42	58	42.0	4560	4	Q13787	APOLIPOPROTEIN B100.	3.50e+01
43	58	42.0	5060	2	Q52345	POLYKETIDE SYNTHASE.	3.50e+01
44	58	42.0	5069	2	Q52789	RFAMYCIN POLYKETIDE S	3.50e+01
45	58	42.0	5082	2	P59514	KIFANOCIN POLYKETIDE SYNTH	3.50e+01

ALIGNMENTS

```

RESULT 1
ID   OS9003          PRELIMINARY;      PRT;      860 AA.
AC   OS9003;
DT   01-AUG-1998   (TREMBLREL. 07, CREATED)
DT   01-AUG-1998   (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-JAN-1999   (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   860AA LONG HYPOTHETICAL PROTEIN.
GN   PH1302.
OS   PYROCOCOCCUS HORIKOSHII.
OC   ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCOCCUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-OT3;
RX   MEDLINE; 98344137.
RA   KAWABAYASHI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA   YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOVAMA A., NAGAI Y.,
RA   SAKAI M., OSGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA   FUNAHASHI T., TANAKA T., KUDOH K., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA   AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA   KIKUCHI H.,
RT   *Complete Sequence and Gene Organization of the Genome of a
RT   Hyper-thermophilic Archaeobacterium, Pyrococcus horikoshii OT3.*;
RL   DNA RES. 5:55-76(1998).
DR   EMBL; AP000006; D1031349; -.
SQ   SEQUENCE 860 AA; 95455 MW; B336F868 CRC32;

Query Match          55.8%; Score 77; DB 1; Length 860;
Best Local Similarity 29.4%; Pred. No. 2.01e-02;
Matches 5; Conservative 9; Mismatches 3; Indels 0; Gaps 0

Db 1 MSKFLTDFSNAFIRVAF 17
    | |||:::|
QY 1 MEKFMAEFGGYVQTF 17

RESULT 2
ID   O74624          PRELIMINARY;      PRT;      364 AA.
AC   O74624;
DT   01-NOV-1998   (TREMBLREL. 08, CREATED)
DT   01-NOV-1998   (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998   (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MANNOSE-1-PHOSPHATE GUANYLYLTRANSFERASE (EC 2.7.7.13).
GN   MPGL.
OS   HYPOCREA JECORINA.

```

```
CC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
DR HYPOCREALES; HYPOCREACEAE; HYPOCREA.
DR [1]
RN SEQUENCE FROM N.A.
RC STRAIN-RUT C-30;
RX MEDLINE: 98309839.
RA KRUSZESKA J.S., SALOHEIMO M., PENTTILA M., PALAMARCZYK G.;
RT "Isolation of a Trichoderma reesei cDNA encoding GTP:
RT alfa-d-mannose-1-phosphate guanylttransferase involved in early steps
RT of protein glycosylation.";
RL CURR. GENET. 33:445-450(1998).
DR EMBL: U99991; G3323397; -.
KW TRANSFERASE; NUCLEOTIDYLTRANSFERASE.
SQ SEQUENCE 364 AA; 40284 MW; 95D4573B CRC32;

Query Match 50.0%; Score 69; DB 3; Length 364;
Best Local Similarity 44.4%; Pred. No. 5.40e-01;
Matches 8; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

Db 60 MEKFLAEYEKYNINIEF 77
QY 1 MEKFLAEFGGY-VQTPF 17

RESULT 3
ID O60074 PRELIMINARY; PRT; 178 AA.
AC O60074;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 19.4 KD PROTEIN.
GN SPBC13G1.12.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972H;
RA LYNE M., WOOD V., RAJANDREAM M.A., BARRELL B.G., BADCOCK K.,
RA CHURCHER C.M.;
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AL022600; E1287787; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 178 AA; 19432 MW; F002D830 CRC32;

Query Match 49.3%; Score 68; DB 3; Length 178;
Best Local Similarity 47.1%; Pred. No. 8.03e-01;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 93 MDKFEAOFDDVNYQTGY 109
QY 1 MEKFLAEFGGY-VQTPF 17

RESULT 4
ID Q42701 PRELIMINARY; PRT; 516 AA.
AC Q42701;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450 (EC 1.14.14.1) (FRAGMENT).
GN CYP72C.
OS CATHARANTHUS ROSEUS (ROSY PERIWINKLE) (MADAGASCAR PERIWINKLE).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; GENTIANACEAE; GENTIANACEAE; APOCYNACEAE; CATHARANTHUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CP3;
RA MANGOLD U., EICHEL J., BATSCHAUER A., LANZ T., KAISER T.,
RA SPANGENBERG G., WERCK-REICHART D., SCHROEDER J.;
RL PLANT SCI. 96:129-136(1994).

-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC EMBL: L19075; G404690; -.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
DR PFAM: PF00067; P450; 1.
DR MENDEL; 11605; CATO:1113;14.
KW OXIDOREDUCTASE; MONOOXYGENASE; ELECTRON TRANSPORT; MEMBRANE; HEME.
FT NON_TER 1
FT BINDING 459 459 HEME (BY SIMILARITY).
SQ SEQUENCE 516 AA; 59720 MW; AD63FB23 CRC32;

Query Match 47.8%; Score 66; DB 10; Length 516;
Best Local Similarity 29.4%; Pred. No. 1.76e+00;
Matches 5; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

Db 477 LORFKFDVAPSYVHAPE 493
QY 1 MEKFLAEFGGY-VQTPF 17

RESULT 5
ID O80784 PRELIMINARY; PRT; 2301 AA.
AC O80784;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE F13P17.19 PROTEIN.
GN F13P17.19.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA SIKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA SOMERVILLE C.R., VENTER J.C.;
RL "Arabidopsis thaliana chromosome II BAC F13P17 genomic sequence.";
DR EMBL: AC004481; G3337366; -.
SQ SEQUENCE 2301 AA; 253398 MW; PB931E6C CRC32;

Query Match 47.1%; Score 65; DB 10; Length 2301;
Best Local Similarity 46.7%; Pred. No. 2.59e+00;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 977 NFLGRFGAGYVSDTF 991
QY 3 KFMFLAEFGGY-VQTPF 17

RESULT 6
ID O67454 PRELIMINARY; PRT; 273 AA.
AC O67454;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 32.9 KD PROTEIN.
GN AQ1477.
OS AQUIFEX AEOLICUS.
OC BACTERIA; AQUIFICALES; AQUIFICACEAE; AQUIFEX.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RX MEDLINE: 98196666.
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL NATURE 392:353-358(1998).
RN [2]
RP SEQUENCE FROM N.A.
```

RC STRAIN-VF5;
RA DECKERT G.; WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBEEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AE000742; G2983869; -
KW HYPOTHETICAL PROTEIN
SQ SEQUENCE 273 AA; 32949 MW; 4B858A2F CRC32;

Query Match 46.4%; Score 64; DB 2; Length 273;
Best Local Similarity 38.5%; Pred. No. 3.80e+00;
Matches 5; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Db 82 QRLDERVEGYID 94
:::|:|:|:|:
QY 2 KEFMAEFGGYVQ 14

RESULT 7
ID Q37744 PRELIMINARY; PRT; 420 AA.
AC Q37744;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PHOSPHOGLYCERATE KINASE (EC 2.7.2.3).
GN CPK-8.
OS TRYPAOSOMA BRUCEI BRUCEI.
OG MITOCHONDRION.
OC EUKARYOTA; EULENOZOA; KINETOPLASTIDA; TRYPAOSOMATIDAE; TRYPAOSOMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EATRO 164;
RX MEDLINE; 91101667.
RA ALEXANDER K.A., PARAIL A.C., PARSONS M.;
RT "An allele of Trypanosoma brucei cytoplasmic phosphoglycerate kinase
is a mosaic of other alleles and genes.";
RL MOL. BIOCHEM. PARASITOL. 42:293-296(1990).
CC -1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE - ADP +
3-PHOSPHO-D-GLYCEROYL PHOSPHATE.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
DR PROSITE; PS00111; PGLYCERATE_KINASE; 1.
DR PFAM; PF00162; PGK; 1.
KW TRANSFERASE; MITOCHONDRION; KINASE; GLYCOLYSIS.
SQ SEQUENCE 420 AA; 45243 MW; 0A2C78BA CRC32;

Query Match 46.4%; Score 64; DB 8; Length 420;
Best Local Similarity 35.3%; Pred. No. 3.80e+00;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 151 MAKILASYGDVYISDAF 167
|:|:|:|:|:|:
QY 1 MEKFAEFGGYVQTPF 17

RESULT 8
ID Q37745 PRELIMINARY; PRT; 440 AA.
AC Q37745;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PHOSPHOGLYCERATE KINASE (EC 2.7.2.3).
OS TRYPAOSOMA BRUCEI BRUCEI.
OG MITOCHONDRION.
OC EUKARYOTA; EULENOZOA; KINETOPLASTIDA; TRYPAOSOMATIDAE; TRYPAOSOMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EATRO 164;
RX MEDLINE; 91101667.
RA ALEXANDER K.A., PARAIL A.C., PARSONS M.;
RT "An allele of Trypanosoma brucei cytoplasmic phosphoglycerate kinase
is a mosaic of other alleles and genes.";
RL MOL. BIOCHEM. PARASITOL. 42:293-296(1990).
CC -1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE - ADP +
3-PHOSPHO-D-GLYCEROYL PHOSPHATE.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
DR PROSITE; PS00111; PGLYCERATE_KINASE; 1.
DR PFAM; PF00162; PGK; 1.
KW TRANSFERASE; MITOCHONDRION; KINASE; GLYCOLYSIS.
SQ SEQUENCE 440 AA; 45243 MW; 0A2C78BA CRC32;

Query Match 46.4%; Score 64; DB 8; Length 509;
Best Local Similarity 35.3%; Pred. No. 3.80e+00;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 240 MAKILASYGDVYISDAF 256
|:|:|:|:|:|:
QY 1 MEKFAEFGGYVQTPF 17

RESULT 10
ID Q42700 PRELIMINARY; PRT; 524 AA.
AC Q42700;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450 (EC 1.14.14.1).
GN CYP72B.
OS CATHARANTHUS ROSEUS (ROSY PERIWINKLE) (MADAGASCAR PERIWINKLE).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; GENTIANAE; GENTIANACEAE; GENTIANALES; APOCYNACEAE; CATHARANTHUS.
RN [1]
RP SEQUENCE FROM N.A.

RL MOL. BIOCHEM. PARASITOL. 42:293-296(1990).
CC -1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE - ADP +
3-PHOSPHO-D-GLYCEROYL PHOSPHATE.
CC -1- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
DR EMBL: M33775; G343585; -
DR PROSITE; PS00111; PGLYCERATE_KINASE; 1.
DR PFAM; PF00162; PGK; 1.
KW TRANSFERASE; MITOCHONDRION; KINASE; GLYCOLYSIS.
SQ SEQUENCE 440 AA; 47231 MW; 09EF6813 CRC32;

Query Match 46.4%; Score 64; DB 8; Length 440;
Best Local Similarity 35.3%; Pred. No. 3.80e+00;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 151 MAKILASYGDVYISDAF 167
|:|:|:|:|:|:
QY 1 MEKFAEFGGYVQTPF 17

RESULT 9
ID Q37743 PRELIMINARY; PRT; 509 AA.
AC Q37743;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PHOSPHOGLYCERATE KINASE (EC 2.7.2.3).
GN P56.
OS TRYPAOSOMA BRUCEI BRUCEI.
OG MITOCHONDRION.
OC EUKARYOTA; EULENOZOA; KINETOPLASTIDA; TRYPAOSOMATIDAE; TRYPAOSOMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EATRO 164;
RX MEDLINE; 91304507.
RA ALEXANDER K.A., PARSONS M.;
RT "A phosphoglycerate kinase-like molecule localized to glycosomal
microbodies: evidence that the topogenic signal is not at the
C-terminus.";
RL MOL. BIOCHEM. PARASITOL. 46:1-10(1991).
CC -1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE - ADP +
3-PHOSPHO-D-GLYCEROYL PHOSPHATE.
CC -1- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
DR EMBL: M37784; G295367; -
DR PROSITE; PS00111; PGLYCERATE_KINASE; 1.
DR PFAM; PF00162; PGK; 1.
KW MITOCHONDRION; TRANSFERASE; KINASE; GLYCOLYSIS.
SQ SEQUENCE 509 AA; 55747 MW; AF7895A4 CRC32;

Query Match 46.4%; Score 64; DB 8; Length 509;
Best Local Similarity 35.3%; Pred. No. 3.80e+00;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 240 MAKILASYGDVYISDAF 256
|:|:|:|:|:|:
QY 1 MEKFAEFGGYVQTPF 17

RESULT 10
ID Q42700 PRELIMINARY; PRT; 524 AA.
AC Q42700;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450 (EC 1.14.14.1).
GN CYP72B.
OS CATHARANTHUS ROSEUS (ROSY PERIWINKLE) (MADAGASCAR PERIWINKLE).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; GENTIANAE; GENTIANACEAE; GENTIANALES; APOCYNACEAE; CATHARANTHUS.
RN [1]
RP SEQUENCE FROM N.A.

Db 264 KFNLFSHNTVQFQF 278
 Qy 3 KFMFEFGGYVQTFP 17

RESULT 12
 ID 084013 PRELIMINARY; PRT: 455 AA.
 AC 084013;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE ACYLTRANSFERASE.
 GN HTRB.
 OS CHLAMYDIA TRACHOMATIS.
 OC BACTERIA; CHLAMYDIALES; CHLAMYDIACEAE; CHLAMYDIA.
 [1]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UM-3/CX;
 RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
 RA MITCHELL W.P., OLINGER L., TATSOV R.L., ZHAO Q., KOONIN E.V.,
 RA DAVIS R.W.;
 RT "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
 RT Chlamydia trachomatis.";
 RN SCIENCE 0:0-0(1998).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UM-3/CX;
 RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
 RA MITCHELL W.P., OLINGER L., TATSOV R.L., ZHAO Q., KOONIN E.V.,
 RA DAVIS R.W.;
 RT SUBMITTED (MAY-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
 DR EMBL; AE001275; G3328398; -;
 KW TRANSFERASE; ACYLTRANSFERASE.
 SQ SEQUENCE 455 AA; 52058 MW; 18AB7CEE CRC32;

Query Match 45.7%; Score 63; DB 2; Length 455;
 Best Local Similarity 29.4%; Pred. No. 5,56e+00;
 Matches 5; Conservative 9; Mismatches 3; Indels 0; Gaps

Db 343 LQRELFIEGFEYADASL 359
 Qy 1 MEKFMFEFGGYVQTFP 17

RESULT 13
 ID 053818 PRELIMINARY; PRT: 645 AA.
 AC 053818;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE PPE-FAMILY PROTEIN.
 GN MYO41.29.
 OS MYCOBACTERIUM TUBERCULOSIS.
 OC BACTERIA; FRIMTICUTES; ACTINOBACTERIA; ACTINOBACTERIACEAE;
 OC ACTINOMYCETALES; CORYNEBACTERIACEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
 [1]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RA HAWLIN N., CHURCHER C.M.;
 RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DDBJ DATA BANKS.
 [2]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RA COLE S.T., PARKHILL J., BARRELL B.G., RAJANDREAN M.A.;
 RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
 [3]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE: 96181548
 RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
 RA COLE S.T.;
 RT "An integrated map of the genome of the tubercle bacillus.

RT Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
 RT leprae.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).
 DR EMBL: AL021958; E1253293; ..
 SQ SEQUENCE 645 AA; 62693 MW; 04F5F7F1 CRC32;

Query Match 45.7%; Score 63; DB 2; Length 645;
 Best Local Similarity 46.7%; Pred. No. 5.56e+00;
 Matches 7; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

Db 592 FMGGFGNTGFLQSGF 606

Qy 4 FMAEFGQ-GYVQTPF 17

RESULT 14

ID O48232 PRELIMINARY; PRT; 1215 AA.

AC O48232;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE DNA FOR SEROTYPE B CAPSULATION LOCUS.

OS HAEMOPHILUS INFLUENZAE.

OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;

OC HAEMOPHILUS.

RN [1]

RC SEQUENCE FROM N.A.

RP STRAIN-RM135;

RX MEDLINE: 95272382.

RA VAN ELDRE J., BROPHY L., LOYNDS B., CELIS P., KROLL J.S., MOXON E.R.,

RA HANCOCK I., CARMAN S.;

RT "Region II of the Haemophilus influenzae type be capsulation locus is

RT involved in serotype-specific polysaccharide synthesis.";

RL MOL. MICROBIOL. 15:107-118(1995).

DR EMBL: X78559; G471236; ..

SQ SEQUENCE 1215 AA; 142323 MW; DF50121E CRC32;

Query Match 45.7%; Score 63; DB 2; Length 1215;

Best Local Similarity 52.6%; Pred. No. 5.56e+00;

Matches 10; Conservative 2; Mismatches 5; Indels 2; Gaps 2;

Db 811 KEKFAFVEFGVGRGTDTF 829

Qy 1 MEKFAEF-GQGYVQTPF 17

RESULT 15

ID O27457 PRELIMINARY; PRT; 191 AA.

AC O27457;

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)

DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE FUCULOSE-1-PHOSPHATE ALDOLASE.

GN MTH1406.

OS METHANOBACTERIUM THERMOAUTOTROPHICUM.

OC ARCHAEA; EURYARCHAOTA; METHANOBACTERIALES; METHANOBACTERIACEAE;

OC METHANOBACTERIUM.

RN [1]

RC SEQUENCE FROM N.A.

RP STRAIN-DELTA H;

RX MEDLINE: 98037514.

RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,

RA ALDRIDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,

RA HARRISON D., HOANG L., KEAGLE P., LUMM W., POTHIER B., QIU D.,

RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,

RA JIWANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PRABHAKAR S.,

RA MDOUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,

RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;

RT "Complete genome sequence of Methanobacterium thermoautotrophicum

RT deltaH: functional analysis and comparative genomics.";

RL J. BACTERIOL. 179:7135-7155(1997).

DR EMBL: AE000903; G2622518; ..

DR PFAM: PF00596; Aldolase-II; 1.

SQ SEQUENCE 191 AA; 20702 MW; BC8B78DC CRC32;

Query Match 44.9%; Score 62; DB 1; Length 191;
 Best Local Similarity 44.4%; Pred. No. 8.10e+00;
 Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 2;

Db 111 MEGFMG-LGRGYIPMPY 127

Qy 1 MEKFAEFGQGYVQ-TPF 17

Search completed: Wed Sep 1 16:08:14 1999
 Job time : 21 secs.

W P E R E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:11:25 1999; MasPar time 5.59 Seconds

Tabular output not generated. 213.200 Million cell updates/sec

Title: >PCT-US99-13024-2

Description: (1-56) from PCTUS9913024.pep (4 of 12)

Perfect Score: 402

Sequence: 1 MEKFMAEFGQGVQTPFLSE.....STAGPSYVKFQDNVPGSQTF 56

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 25.232; Variance 100.718; scale 0.251

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	300	74.6	65.19	R97371	Phage T4 ORF gene pr	1.93e-20
2	81	20.1	3011	12 R66995	Hepatitis C virus gen	1.51e+01
3	81	20.1	3011	8 R40120	HCV genomic amino aci	1.51e+01
4	80	19.9	170	1 P90150	Sequence of hepatitis	1.82e+01
5	80	19.9	170	1 P92033	Sequence encoded in t	1.82e+01
6	80	19.9	411	16 R90934	HCV NS5 domain antige	1.82e+01
7	80	19.9	499	38 W67010	HCV non-structural pr	1.82e+01
8	80	19.9	516	6 R33441	HCV NS5 encoded by ph	1.82e+01
9	80	19.9	516	7 R33633	HCV CKS-NS5E fusion p	1.82e+01
10	80	19.9	516	6 R33575	HCV CKS-NS5E recombin	1.82e+01
11	80	19.9	516	6 R33595	HCV CKS-NS5E recombin	1.82e+01
12	80	19.9	516	6 R21566	HCV CKS-NS5E - pHCV-4	1.82e+01
13	80	19.9	798	7 R33630	HCV CKS-NS5 EF fusion	1.82e+01
14	80	19.9	1766	1 P92041	Sequence encoded in t	1.82e+01
15	80	19.9	1786	1 P90158	Protein sequence of h	1.82e+01
16	80	19.9	2261	1 P90164	Peptide encoded by co	1.82e+01

17	80	19.9	2301	1	P92047	Sequence encoded in t
18	80	19.9	2435	5	R25135	HCV polypeptide 1.
19	80	19.9	2436	1	P92050	Sequence encoded in t
20	80	19.9	2436	5	R28582	HCV amino acid sequen
21	80	19.9	2462	1	P90288	Peptide encoded by co
22	80	19.9	2772	2	R08123	Hepatitis C virus pol
23	80	19.9	2816	7	R34009	HCV-1 polypeptide.
24	80	19.9	2894	5	R24440	Composite HCV HC-J1/C
25	80	19.9	2894	13	R70230	Composite hepatitis C
26	80	19.9	2955	2	R08124	Hepatitis C virus put
27	80	19.9	3011	26	W34480	HCV polypeptide.
28	80	19.9	3011	4	R21519	Completed HCV sequence
29	80	19.9	3011	28	W40038	HCV polypeptide.
30	80	19.9	3011	16	R09031	Hepatitis C virus pol
31	80	19.9	3011	6	R31621	Hepatitis C virus (HC
32	79	19.7	458	11	R56975	Variant vitamin D bin
33	79	19.7	458	4	R22278	Human gp.-specific co
34	79	19.7	458	23	W10561	Vitamin D3-binding pr
35	79	19.7	458	11	R56976	Variant vitamin D bin
36	79	19.7	458	4	R22279	Human gp.-specific co
37	79	19.7	1464	10	R55529	Human NMDA R2A recept
38	78	19.4	3011	36	W77397	Hepatitis C virus H77
39	78	19.4	3011	4	R22154	NANBV Hutch c59 isola
40	78	19.4	3011	8	R40119	HCV genomic amino aci
41	78	19.4	3011	14	R79232	HCV sequence.
42	78	19.4	3011	36	W77398	Hepatitis C virus-H C
43	77	19.2	135	35	W76234	Bacterial periplasmic
44	77	19.2	1464	18	R80970	Human excitatory amin
45	77	19.2	1464	38	W85576	Human N-methyl-D-aspa

ALIGNMENTS

RESULT 1
ID R97371 standard; Protein; 65 AA.
AC R97371;
DE 07-JAN-1997 (first entry)
DT Phage T4 ORF gene product. gp34.
KW filter; tail fibre protein; nanotechnology; nano-structure;
OS Bacteriophage T4.
OS Bacteriophage T4.
PD 25-APR-1996.
PF 13-OCT-1995; U13023.
PR 13-OCT-1994; US-322760.
PA (GOLD/) GOLDBERG E B.
PI Goldberg EB;
DR WPI; 96-221942/22.
DR N-PSDB; T29053.
PT New proteins derived from T4 phage tail fibre proteins - that can self assemble into nano-structure(s), useful as filters etc, also corresponding DNA
PS Claim 7; Fig 7; 83pp; English.
CC A protein (R97371) of unspecified function is the product of open reading frame x of the tail fibre protein gene region (see also T29053) of phage T4. This gene region also includes open reading frames for tail fibre proteins (see also R97370 and R97372-74).
CC Tail fibre proteins (native or modified) can be produced in large quantities in microbial cells and used as bulding blocks of strong, stable nanostructures.
SQ Sequence 65 AA;

Query Match 74.6%; Score 300; DB 19; Length 65;

Best Local Similarity 82.1%; Pred.No. 1.93e-20;

Matches 46; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Db 1 mekfmawetwricpnaalisesnsvrykisiagscplstagspyvkfqnpgvsqtf 56

QY 1 MEKFMAEFGQGVQTPFLSESNVRYKISAGSCPLSTAGSPYVKFQDNVPGSQTF 56

RESULT 2

ID R66995 standard; Protein; 3011 AA.

[illegible]

RESULT 6

ID R90934 standard; Protein; 411 AA.

AC R90934;

DT 15-MAY-1996 (first entry)

DE HCV NS5 domain antigen.

KW Non-A non-B hepatitis virus: NANBHV; HCV; antigen; detection;

KW diagnosis; antibodies.

OS Hepatitis C virus.

PN EP-693687-A1.

PD 24-JAN-1996.

PF 03-APR-1991; 114016.

PR 04-APR-1990; US-504352.

PA (CHIR) CHIRON CORP.

PI Choo Q, Houghton M, Kuo G;

DR WPI; 96-117956/13.

PT Combinations of synthetic Hepatitis C Virus antigens - provide more effective diagnosis of Non-A, Non-B Hepatitis

PS Claim 8; Fig 1(A-Y); 53pp; English

CC The combination comprises an HCV antigen from the C domain (pref. C22 - R90936) and at least one HCV antigen from the NS3 (pref. C33c C22 - R90932), NS4 (pref. C100 - R90933), S (pref. S2 - R90935) or NS5 (R90934) domain.

CC The antigens may in the form of a fusion protein, a simple physical mixture, or the individual antigens commonly bound to a solid matrix.

CC They are pref. prepd. by recombinant DNA techniques (primers are given in T12711-T12716), but can be synthesised or isolated from HCV using affinity chromatography.

CC Sequence 411 AA;

SQ

Query Match 19.9%; Score 80; DB 16; Length 411;
Best Local Similarity 37.8%; Pred. No. 1.82e+01;
Matches 17; Conservative 11; Mismatches 13; Indels 4; Gaps 4;

```
Db      63 vpspeffteIdgvrIh-rfappckpllreevsfrvlgIheypvsgq 106  
        | : | : | : | : | : | : | : | : | : | : | : |  
Qy     13 VQTP-FLSESNVRYKISIAAGSC-PLSTAGPSY-VKFQDNPVGSQ 54
```

RESULT	7	
ID	W67010 standard; protein; 499 AA.	
AC	W67010;	
DT	02-MAR-1999 (first entry)	
DE	HCV non-structural protein NS4.	
KW	Hepatitis C virus; HCV; nucleocapsid; core protein; analogue; antibody;	
KW	non-structural protein; thioamide bond; peptide bond.	
OS	Hepatitis C virus.	
PN	J10226698-A.	
PD	25-AUG-1998.	
PF	19-FEB-1997; 034702.	
PR	19-FEB-1997; JP-034702.	
PA	(KYOW) KYOWA MEDEX KK.	
DR	WPI: 98-515103/44.	
PT	Determination of antibody in sample - uses peptide analog absorbed	
PT	or chemically bound on carrier as antigen	
PS	Disclosure; Page 5; 13pp; Japanese.	
CC	This sequence represents the Hepatitis C virus (HCV) non-structural (NS4	
CC	protein. The invention relates to peptide analogues derived from HCV	
CC	proteins, e.g. W67417-W67426, which can be used for the determination	
CC	of anti-HCV antibodies in a sample. Preferably the peptide analogues	
CC	contain one or more thioamide peptide bonds where at least one oxygen	
CC	atom of the peptide bond is replaced by sulphur atom. The peptide	
CC	analogues can be adsorbed or chemically bound to a carrier.	
SQ	Sequence 499 AA;	

```

Query Match      19.98; Score 80; DB 38; Length 499;
Best Local Similarity 37.84; pred. No. 1.82e+01;
Matches 17; Conservative 11; Mismatches 13; Indels 4; Gaps 4;

Db 48 vpspeffteidgrllh-rtappckpllleevsfrvlghyevqsg 91
QY 13 VQTP-FLSESNSVRYIKISAGSE-PLSTAGPSY-VKFDPNQVSG 54

```

RESULT	ID	R33441	R33441	8
AC	AC	30-JUN	30-JUN	
DE	DE	HCV NS5	HCV NS5	
KW	KW	Monoclon	Monoclon	
OS	OS	Hepatitis	Hepatitis	
PN	PN	W09304	W09304	
PD	PD	04-MAR	04-MAR	
PF	PF	21-AUG	21-AUG	
PR	PR	21-AUG	21-AUG	
PA	PA	(ABBO	(ABBO	
PI	PI	Dailey	Dailey	
DR	DR	WPI; 9	WPI; 9	
PT	PT	Monoclon	Monoclon	
PT	PT	diagnosis	diagnosis	
PS	PS	Disclo	Disclo	
CC	CC	The sec	The sec	
CC	CC	virus l	virus l	
CC	CC	which	which	
CC	CC	cell l	cell l	
CC	CC	See a	See a	
CC	CC	Sequenc	Sequenc	
CC	CC			

Query Match		
Best Local		
Matches		
Db	433	vff
QY	13	vc

	9
RESULT	R33633
ID	AD
AC	R33633
DE	13-JUL-78
DT	HCV CRK
KW	Hepatitis
KW	non-st
KW	CYP-CMP
KW	CTP-CMP
KW	Immunop
KW	Hepatitis
PN	WO9304
PD	04-MAR
PF	21-AUG
PF	21-AUG
PA	(ABBO)
PI	Dalley
PI	WPI
DR	N-P5DB
DR	New re
PT	detect
PT	Example
PS	Eight c
CC	HCV gen
CC	fragment
CC	sequenc
CC	fusion
CC	nine am
CC	acids r
CC	recomb
CC	with hu
SQ	Sequenc

Query Match		
Best Local	433	vp
Matches		
		13
	Qy	vo

```

CC HCV.
SQ Sequence 516 AA:

Query Match 19.9% Score 80; DB 6; Length 516;
Best Local Similarity 37.8%; Pred. No. 1.82e+01;
Matches 17; Conservative 11; Mismatches 13; Indels 4; Gaps 4;

Db 433 vpspeffteidgvrhl-rfappckpdlreevsfrvlgheypvqsq 476
QY 13 VQTP-FLSENSVRYKISIAGSC-PLSTAGPSY-VKFDQNPVGSQ 54
| | | | | : | | | | | : | | | | | : | | | | |
| | | | | : | | | | | : | | | | | : | | | | |

RESULT 12
ID R21566 standard; Protein; 516 AA.
AC R21566;
DE 09-JUN-1992 (first entry)
DT HCV CKS-NS5E - pHCV-45.
KW Hepatitis C virus; antigen; diagnosis; inhibitor; CMP-KDO synthase;
KW CKS; HCV CKS-NS5E; NANBHV.
FH Key Location/Qualifiers
FT Peptide 1..239
FT /label= CKS
FT region 240..248
FT /label= linker
FT /note= "see CC"
FT peptide 249..508
FT /label= NS5-E
FT /note= "HCV region NS4/NS5, amino acids 1932-2191"
FT region 509..516
FT /label= linker
FT /note= "see CC"
FT EP-472207-A.
PD 26-FEB-1992.
PF 23-AUG-1991; 114161.
PR 24-AUG-1990; US-572822.
PR 07-NOV-1990; US-614069.
PA (ABBO ) ABBOTT LABORATORIES.
PI Devare SG, Desai SM, Casey JM, Dawson GJ, Lesniewski RR;
PI Dailey SH, Gutierrez RA, Stewart JL;
DR WPI: 92-066430/09.
DR N-PSDB; Q21679.
PT Recombinant hepatitis C virus antigens - produced as fusion
PT proteins and representing distinct antigenic regions of the HCV
PT genome
PS Disclosure: Fig 26, Page 61-63; 115pp; English.
CC Note: according to the diagrammatic illustration of this sequence,
CC both linkers comprise 9 amino acids.
CC The polypeptide (mol.wt. 55 kD) represents a distinct antigenic
CC region of the HCV genome and can be used for the detection of
CC antibodies and antigens for early diagnosis of HCV infection.
CC The polypeptide can also be used to develop specific inhibitors of
CC viral replication and for therapeutic purposes.
SQ Sequence 516 AA;

Query Match 19.9% Score 80; DB 4; Length 516;
Best Local Similarity 37.8%; Pred. No. 1.82e+01;
Matches 17; Conservative 11; Mismatches 13; Indels 4; Gaps 4;

Db 433 vpspeffteidgvrhl-rfappckpdlreevsfrvlgheypvqsq 476
QY 13 VQTP-FLSENSVRYKISIAGSC-PLSTAGPSY-VKFDQNPVGSQ 54
| | | | | : | | | | | : | | | | | : | | | | |
| | | | | : | | | | | : | | | | | : | | | | |

RESULT 13
ID R33630 standard; protein; 798 AA.
AC R33630;
DE 13-JUL-1993 (first entry)
DT HCV CKS-NS5 EF fusion protein pHCV59.
KW Hepatitis C virus; NANBH; non-A, non-B hepatitis; CMP-KDO synthetase;
KW non-structural protein; pHCV-59; diagnosis; immunosay;
KW CTP;CMP-3-deoxymanno-octulosonate cytidyl transferase.
OS Hepatitis C Virus.
PN WO9304089-A

```



```
##status      preliminary
##molecule_type DNA
##residues    1-869 #label JOR
##cross-references GB:M55661; NID:g145507; PID:g145510
##experimental_source enterotoxigenic strain, CFA/I-ST plasmid NTP113
##note        sequence extracted from NCBI backbone (NCBIN:108960,
               NCBIP:108971)

GENETICS
#gene         cfaC
#genome       plasmid
SUMMARY
#length 869 #molecular-weight 97830 #checksum 9755

Query Match      21.1%; Score 85; DB 2; Length 869;
Best Local Similarity 32.4%; Pred. No. 2.31e-01;
Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;

Db 158 AFTQSQTINLSDGKYKRLISGNSALGITDTSY 191
      ::::: :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
QY 11 GYVOTPFLSENSVRYK-ISIAGSCPLSTAGPSY 43

RESULT 3 ~
ENTRY
TITLE      I38375 #type fragment
ALTERNATE_NAMES tyrosine kinase - human (fragment)
ORGANISM     #formal_name Homo sapiens #common_name man
DATE         29-May-1998 #sequence_revision 29-May-1998 #text_change
            10-Jul-1998

ACCESSIONS   I38375
REFERENCE    I38372
#authors     Haire, R.N.; Ohta, Y.; Lewis, J.E.; Fu, S.M.; Kroisel, P.;
            Litman, G.W.
#journal     Hum. Mol. Genet. (1994) 3:897-901
#title       TXK, a novel human tyrosine kinase expressed in T cells
            shares sequence identity with Tec family kinases and maps
            to 4p12
#cross-references MUID:95038742
#accession   I38375
##status     preliminary; translated from GB/EMBL/DDBJ
##molecule_type DNA
##residues  1-81 #label RES
##cross-references EMBL:U07794; NID:g508219; PID:g508224

GENETICS
#gene        GDB:TXK
#map_position 4p12-4p12
#introns     25/1
CLASSIFICATION #superfamily protein-tyrosine kinase tec; pleckstrin repeat
               homology; protein kinase homology; SH2 homology; SH3
               homology

FEATURE
32-81
SUMMARY      #domain protein kinase homology (fragment) #label KIN
               #length 81 #checksum 7751

Query Match      20.4%; Score 82; DB 2; Length 81;
Best Local Similarity 44.4%; Pred. No. 6.47e-01;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 5 LRPVGLMGSLCPATAGFSYEKWEIDP 31
      ::|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
QY 24 VRYKISAGSCPLSTAGPSVVKFQDNP 50

RESULT 4
ENTRY
TITLE      I84483 #type complete
ALTERNATE_NAMES tyrosine kinase - human
ORGANISM     #formal_name Homo sapiens #common_name man
DATE         29-May-1998 #sequence_revision 29-May-1998 #text_change
            12-Feb-1999

ACCESSIONS   I84483
REFERENCE    I38372
#authors     Haire, R.N.; Ohta, Y.; Lewis, J.E.; Fu, S.M.; Kroisel, P.;
            Litman, G.W.
#journal     Hum. Mol. Genet. (1994) 3:897-901

TXK, a novel human tyrosine kinase expressed in T cells
shares sequence identity with Tec family kinases and maps
to 4p12

#cross-references MUID:95038742
#accession   I84483
##status     preliminary; translated from GB/EMBL/DDBJ
##molecule_type mRNA
##residues  1-527 #label HUA
##cross-references GB:L35268; NID:g623442; PID:g623443
CLASSIFICATION #superfamily protein-tyrosine kinase tec; pleckstrin repeat
               homology; protein kinase homology; SH2 homology; SH3
               homology
               phosphotransferase

KEYWORDS
FEATURE
89-137
150-246
269-527
277-285
SUMMARY      #domain SH3 homology #label SH3\
               #domain SH2 homology #label SH2\
               #domain protein kinase homology #label KIN\
               #region protein kinase ATP-binding motif
               #length 527 #molecular-weight 61108 #checksum 785

Query Match      20.1%; Score 81; DB 2; Length 527;
Best Local Similarity 44.4%; Pred. No. 9.06e-01;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 242 LRPVGLMGSLCPATAGFSYEKWEIDP 268
      ::|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
QY 24 VRYKISAGSCPLSTAGPSVVKFQDNP 50

RESULT 6
ENTRY
TITLE      I49133 #type complete
```


TITLE	Txk - mouse
ORGANISM	#formal_name Mus musculus #common_name house mouse
DATE	02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 12-Feb-1999
ACCESSIONS	I49133
REFERENCE	I49133
#authors	Haire, R.N.; Litman, G.W.
#journal	Mamm. genome (1995) 6:476-480
#title	The murine form of TXK, a novel TBC kinase expressed in thymus maps to chromosome 5.
#cross-references	MUTID:96059536
#accession	I49133
#status	Preliminary; translated from GB/EMBL/DBDJ
#molecule_type	mRNA
#residues	1-527 #label RES
#cross-references	EMB1:U16145; NID:g562124; PID:g562125
CLASSIFICATION	#superfamily protein-tyrosine kinase tcc; pleckstrin repeat homology; protein kinase homology; SH2 homology; SH3 homology
KEYWORDS	ATP
FEATURE	
89-137	#domain SH3 homology #label SH3\
150-246	#domain SH2 homology #label SH2\
269-527	#domain protein kinase homology #label KIN\
277-285	#region protein kinase ATP-binding motif
SUMMARY	#length 527 #molecular-weight 61108 #checksum 785
Query Match	20.1% Score 81; DB 2; Length 527;
Best Local Similarity	44.4%; Pred. No. 9, 06e-01;
Matches	12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
Dd	242 LRYPIGLGSLCPATSGFSYKEWIDP 268 ::: : : : :
Oy	24 VRYKISAGSCPLSTAGSYVKFQDNP 50 :
RESULT	7
ENTRY	S40770 #type complete
TITLE	polyprotein precursor - hepatitis C virus
ORGANISM	#formal_name hepatitis C virus
DATE	20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 18-Sep-1998
ACCESSIONS	S40770; PC1285
REFERENCE	S40770
#authors	Okamoto, H.
#submission	submitted to the EMBL Data Library, March 1992
#accession	S40770
#status	Preliminary
#molecule_type	genomic RNA
#residues	1-3011 #label OKA
#cross-references	EMB1:D10749; NID:g221586; PID:d1002057; PID:g221587
REFERENCE	PC1284
#authors	Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsuda, F.; Miyakawa, Y.; Mayumi, M.
#journal	J. Exp. Med. (1990) 60:167-177
#title	The 5'-terminal sequence of the hepatitis C virus genome.
#accession	PC1285
#molecule_type	genomic RNA
#residues	1-513 #label OK2
#experimental_source	isolate HC-J1
CLASSIFICATION	#superfamily hepatitis C virus genome polyprotein; DEAD/H box helicase homology
KEYWORDS	polyprotein
FEATURE	
1230-1493	#domain DEAD/H box helicase homology #label DEAD\
1230-1237	#region nucleotide-binding motif A (P-loop)
SUMMARY	#length 3011 #molecular-weight 327114 #checksum 5911
Query Match	20.1% Score 81; DB 2; Length 3011;
Best Local Similarity	40.0%; Pred. No. 9, 06e-01;
Matches	18; Conservative 10; Mismatches 13; Indels 4; Gaps 4;
Dd	2116 VPSPFFTELDGVRLH-RFAPPCKPLLREEVSRVGLHDYPVGQS 2159

[illegible]

[illegible]

OY 1 MEKFMAEFGGYYVQTPFLSESNVRYKISIAAGSCPLS 37

Search completed: Wed Sep 1 16:11:07 1999
Job time : 16 secs.

W P S R L H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 1 16:09:43 1999; MasPar time 3.74 Seconds
Tabular output not generated. 422.797 Million cell updates/sec

Title: >PCT-US99-13024-2
Description: (1-56) from PCTUS9913024.pep (4 of 12)
Perfect Score: 402
Sequence: 1 MEKFAEFGGQYVQTFPLE.....STAGPSYVKFQDNVPGSOTF 56

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 36.535; Variance 57.714; scale 0.633

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query %	Match	Length	DB	ID	Description	Pred. No.
1	300	74.6	65	1	Y15A_BPT4			HYPOTHETICAL 7.3 KD PR	4.61e-49
2	85	21.1	869	1	CFAC_ECOLI			CYP11 FIMBRIAL SUBUNIT	6.62e-02
3	82	20.4	527	1	TXK_HUMAN			TYROSINE-PROTEIN KINAS	2.04e-01
4	81	20.1	527	1	TKX_MOUSE			TYROSINE-PROTEIN KINAS	2.96e-01
5	80	19.9	802	1	NAB3_YEAST			NUCLEAR POLYADENYLATED	4.27e-01
6	80	19.9	3011	1	POLG_HCV1			GENOME POLYPROTEIN [CO	4.27e-01
7	79	19.7	472	1	VTDB_MOUSE			VITAMIN D-BINDING PROT	6.14e-01
8	79	19.7	474	1	VTDB_HUMAN			VITAMIN D-BINDING PROT	6.14e-01
9	78	19.4	353	1	YQHT_BACSU			PUTATIVE PEPTIDASE IN	8.81e-01
10	78	19.4	476	1	VTDB_RABIT			VITAMIN D-BINDING PROT	8.81e-01
11	78	19.4	3011	1	POLG_HCVH			GENOME POLYPROTEIN [CO	8.81e-01
12	77	19.2	1464	1	NMEL_MOUSE			GLUTAMATE [NMDA] RECEP	1.26e+00
13	77	19.2	1464	1	NMEL_RAT			GLUTAMATE [NMDA] RECEP	1.26e+00
14	76	18.9	338	1	TCPF_VIBCH			TOXIN CORRELATED FIL	1.79e+00
15	75	18.7	575	1	HENA_INBBO			HEMAGGLUTININ PRECURSO	2.54e+00
16	75	18.7	591	1	EYAL_MOUSE			EYES ABSENT HOMOLOG 1.	2.54e+00
17	75	18.7	592	1	EYAL_HUMAN			EYES ABSENT HOMOLOG 1.	2.54e+00
18	75	18.7	984	1	DPOL_NPVAC			DNA POLYMERASE (EC 2.7	2.54e+00
19	75	18.7	986	1	DPOL_NPVBM			DNA POLYMERASE (EC 2.7	2.54e+00
20	74	18.4	171	1	YC92_PSEDE			HYPOTHETICAL 19.0 KD P	3.60e+00
21	73	18.2	105	1	CUI2_HYACE			FLEXIBLE CUTICLE PROTEI	5.07e+00
22	73	18.2	130	1	RS8_METVA			30S RIBOSOMAL PROTEIN	5.07e+00
23	73	18.2	339	1	PHR_BACFI			DEOXYRIBODIPYRIMIDINE	5.07e+00

24 73 18.2 706 1 CATE_MYCAV CATALASE HP11 (EC 1.11 5.07e+00
25 73 18.2 898 1 RIR1_YEAST RIBONUCLEOSIDE-DIPHOS 5.07e+00
26 73 18.2 1363 1 ILPR_BRLA INSULIN-LIKE PEPTIDE R 5.07e+00
27 73 18.2 1693 1 POLN_HEVMY NON-STRUCTURAL POLIPRO 7.11e+00
28 72 17.9 260 1 COAT_MDV COAT PROTEIN. 7.11e+00
29 72 17.9 282 1 Y265_MYCPN HYPOTHETICAL PROTEIN M 7.11e+00
30 72 17.9 345 1 GBA4_DICD1 GUANINE NUCLEOTIDE-BIN 7.11e+00
31 72 17.9 345 1 HEMA_INBF6 HEMAGGLUTININ (FRAGMEN 7.11e+00
32 72 17.9 347 1 HEMA_INBF9 HEMAGGLUTININ PRECURSO 7.11e+00
33 72 17.9 347 1 HEMA_INBF9 HEMAGGLUTININ PRECURSO 7.11e+00
34 72 17.9 354 1 HEM6_MOUSE COPROPORPHYRINOGEN III 7.11e+00
35 72 17.9 457 1 EMB8_PICGL LATE EMBRYOGENESIS ABU 7.11e+00
36 72 17.9 476 1 VTDB_RAT VITAMIN D-BINDING PROT 7.11e+00
37 72 17.9 631 1 TEC_HUMAN TYROSINE-PROTEIN KINAS 7.11e+00
38 72 17.9 695 1 GFAL_SCHPO PUTATIVE GLUCOSAMINE-- 9.95e+00
39 71 17.7 315 1 TRXB_MYCPN THIOREDUXIN REDUCTASE 9.95e+00
40 71 17.7 318 1 NULM_HORSE NADH-UBIQUINONE OXIDOR 9.95e+00
41 71 17.7 347 1 HEMA_INBF3 HEMAGGLUTININ PRECURSO 9.95e+00
42 71 17.7 588 1 CAR7_CANAL CANDIDAPEPSIN 7 PRECUR 9.95e+00
43 71 17.7 611 1 PES4_YEAST PES4 PROTEIN (DNA POLY 9.95e+00
44 71 17.7 935 1 YJ31_YEAST HYPOTHETICAL 108.4 KD 9.95e+00
45 71 17.7 3412 1 POLG_TBVEVS GENOME POLYPROTEIN [CO 9.95e+00

ALIGNMENTS

RESULT 1
ID Y15A_BPT4 STANDARD; PRT: 65 AA.
AC P39509;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DE HYPOTHETICAL 7.3 KD PROTEIN IN GP34-GP35 INTERGENIC REGION.
GN Y15A OR 34.1.
OS BACTERIOPHAGE T4.
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE;
OC T4-LIKE PHAGES.
RN [1]
RP SEQUENCE FROM N.A.
RA KUTTER E.M.;
RL SUBMITTED (NOV-1994) TO THE SWISS-PROT DATA BANK.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 65 AA; 7334 MW; C2D7CE2D CRC32;

Query Match 74.6%; Score 300; DB 1; Length 65;
Best Local Similarity 82.1%; Pred. No. 4.61e-49;
Matches 46; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Db 1 MEKFAEFGGQYVQTFPLES...STAGPSYVKFQDNVPGSOTF 56
QY 1 MEKFAEFGGQYVQTFPLES...STAGPSYVKFQDNVPGSOTF 56

RESULT 2
ID CFAC_ECOLI STANDARD; PRT: 869 AA.
AC P25733;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DE CFA/I FIMBRIAL SUBUNIT C PRECURSOR (COLONISATION FACTOR ANTIGEN I DE SUBUNIT C).
GN CFAC.
OS ESCHERICHIA COLI.
OC PLASMID NTP513.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ENTEROTOXIGENIC;
RX MEDLINE; 99330163.
RA HAMERS A.M., PEL H.J., WILLSHAW G.A., KUSTERS J.G.,
RA VAN DER ZEIJST B.A.M., GAASTRA W.;
RT *The nucleotide sequence of the first two genes of the CFA/I fimbrial

```
RT operon of human enterotoxigenic Escherichia coli." ;
RL MICROR. PATHOG. 6:297-309(1989).
RN [2]
RX SEQUENCE FROM N.A.
RX MEDLINE: 92329981.
RA JORDI B.J.A.M., WILLSHAW G.A., VAN DER ZEIJST B.A.M., GAASTRA W.;
RT "The complete nucleotide sequence of region 1 of the CFA/I fimbrial
operon of human enterotoxigenic Escherichia coli." ;
RL DNA SEQ. 2:257-263(1992).
CC -!- FUNCTION: MAY SERVE AS ANCHOR FOR THE FIMBRIAE IN THE OUTER
CC MEMBRANE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M55661: G145510: -
KW ANTIGEN; SIGNAL; FIMBRIA; OUTER MEMBRANE; PLASMID.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 869 CFA/I FIMBRIAL SUBUNIT C.
SQ SEQUENCE 869 AA: 97830 MW: 7AF76347 CRC32;
Query Match 21.1%; Score 85; DB 1; Length 869;
Best Local Similarity 32.4%; Pred. No. 6.62e-02;
Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;
Db 158 AFTOSOTINLSGSKYKRLISGNSALGITDTSY 191
QY 11 GYVQTPELSESNVRYK-ISIAGSCLSTAGPSY 43
RESULT 3
ID TXK_HUMAN STANDARD; PRT; 527 AA.
AC P42681; Q14220;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE TXK (EC 2.7.1.112).
GN TXK.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BLOOD;
RX MEDLINE: 95038742.
RA HAIRE R.N., OHTA Y., LEWIS J.E., FU S.M., KROISEL P.M., LITMAN G.W.;
RT "TXK, a novel human tyrosine kinase expressed in T cells shares
sequence identity with Tec family kinases and maps to 4p12." ;
RL HUM. MOL. GENET. 3:897-901(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=BLOOD;
RX MEDLINE: 96197775.
RA OHTA Y., HAIRE R.N., AMEMIYA C.T., LITMAN R.T., TRAGER T., RIESS O.,
RA LITMAN G.W.;
RT "Human Txk: genomic organization, structure and contiguous physical
linkage with the Tec gene." ;
RL ONCOGENE 12:937-942(1996).
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN T CELLS AND SOME MYELOID CELL
CC LINES.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN. BELONGS TO THE BTK SUBFAMILY.
CC -----
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L27071: G684986: -
DR EMBL: U34379: G1161364: -
DR EMBL: U34367: G1161364: JOINED.
DR EMBL: U34368: G1161364: JOINED.
DR EMBL: U34369: G1161364: JOINED.
DR EMBL: U34370: G1161364: JOINED.
DR EMBL: U34371: G1161364: JOINED.
DR EMBL: U34372: G1161364: JOINED.
DR EMBL: U34373: G1161364: JOINED.
DR EMBL: U34374: G1161364: JOINED.
DR EMBL: U34375: G1161364: JOINED.
DR EMBL: U34376: G1161364: JOINED.
DR EMBL: U34377: G1161364: JOINED.
DR EMBL: U34378: G1161364: JOINED.
DR MIM: 600058: -
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS50001; SH2; 1.
DR PROSITE: PS50002; SH3; 1.
DR PFAM: PF00017; SH2; 1.
DR PFAM: PF00018; SH3; 1.
DR PFAM: PF00069; pkinase; 1.
DR HSP: Q06187; LAWV.
KW TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;
KW SH3 DOMAIN; PHOSPHORYLATION.
FT DOMAIN 14 19 POLY-CYS.
FT DOMAIN 68 73 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 82 142 SH3.
FT DOMAIN 150 246 SH2.
FT DOMAIN 271 527 PROTEIN KINASE.
FT NP_BIND 277 285 ATP (BY SIMILARITY).
FT BINDING 299 299 ATP (BY SIMILARITY).
FT ACT_SITE 390 390 BY SIMILARITY.
FT MOD_RES 420 420 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CONFLICT 45 45 H -> R (IN REF. 2).
SQ SEQUENCE 527 AA; 61239 MW; 8DF019E3 CRC32;
Query Match 20.4%; Score 82; DB 1; Length 527;
Best Local Similarity 44.4%; Pred. No. 2.04e-01;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
Db 242 LRYPVGLMGSLPATAGFSYKWEIDP 268
QY 24 VRYKISAGSCPLSTAGPSYVRFQDNP 50
RESULT 4
ID TXK_MOUSE STANDARD; PRT; 527 AA.
AC P42682;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE TXK (EC 2.7.1.112) (PTK-RL-18) (RESTING
DE LYMPHOCYTE KINASE).
GN TXK OR RLK
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=THYMUS;
RX MEDLINE: 96059536.
RA HAIRE R.N., LITMAN G.W.;
RT "The murine form of Txk, a novel TEC kinase expressed in thymus maps
```

to chromosome 5. ";
MAMM. GENOME 6:476-480(1995).
[2]
SEQUENCE FROM N.A.
STRAIN-FVB/N; TISSUE=THYMUS;
SOMMERS C.L., HUANG K., GRINBERG A., CHARLICK D.A., KOZAK C.A.,
RA LOVE P.E.;
SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
[3]
SEQUENCE FROM N.A.
STRAIN=C57BL/6; TISSUE=LIVER;
RA HIGASHITSUJI H., NONOGUCHI K., ARII S., FURUTANI M., KANEKO Y.,
RA NAKAYAMA H., FUJITA J.;
SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
[4]
SEQUENCE FROM N.A.
TISSUE=THYMUS;
MEDLINE: 95130578.
HU Q., DAVIDSON D., SCHWARTZBERG P.L., MACCHIARINI F.,
RA LENARDO M.J., BLUESTONE J.A., MARIS L.A.;
Identification of Rlk, a novel protein tyrosine kinase with
predominant expression in the T cell lineage.";
J. BIOL. CHEM. 270:1928-1934(1995).
CC -I- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -I- TISSUE SPECIFICITY: EXPRESSED IN EARLY THYMOCYTES, T CELLS AND
CC MAST CELLS.
CC -I- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -I- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN. BELONGS TO THE BTK SUBFAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: U16145; G562125; -
DR EMBL: U19607; G643065; -
DR EMBL: D43963; G604884; -
DR EMBL: L35268; G623443; -
DR MGD; MGI:102960; TXK.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS50001; SH2; 1.
DR PROSITE; PS50002; SH3; 1.
DR PFAM; PF00017; SH2; 1.
DR PFAM; PF00018; SH3; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSP; Q06187; IAWW
KW TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;
KW SH3 DOMAIN; PHOSPHORYLATION.
FT DOMAIN 14 20 POLY-CYS.
FT FT 82 142 SH2.
FT FT 150 246 SH3.
FT FT 271 527 PROTEIN KINASE.
FT FT 277 285 ATP (BY SIMILARITY).
FT BINDING 299 299 ATP (BY SIMILARITY).
FT ACT_SITE 390 390 BY SIMILARITY.
FT MOD_RES 420 420 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CONFLICT 3 4 LS -> SF (IN REF. 3).
FT CONFLICT 6 6 Y -> D (IN REF. 3).
FT CONFLICT 272 272 A -> T (IN REF. 3).
FT CONFLICT 497 497 R -> S (IN REF. 3).
SQ SEQUENCE 527 AA; 61108 MW; 5B39DA78 CRC32;

Query Match 20.1%; Score 81; DB 1; Length 527;
Best Local Similarity 44.4%; Pred. No. 2.96e-01;

```

Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 242 LRYPIGLGSLPATSGFSYKWEIDP 268
      :| | :| | :| | :| | :| | :| | :| | :| |
Qy 24 VRYKISAGSCLPLSTAGPSYVKFDNP 50

RESULT 5
ID NAB3_YEAST STANDARD; PRT; 802 AA.
AC P38996;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE NUCLEAR POLYADENYLATED RNA-BINDING PROTEIN NAB3.
GN NAB3 OR YPL190C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
[1]
RN SEQUENCE FROM N.A.
RA WILSON S.M., OBERDORF A.M., DATAR K.V., SWEDLOW J.R., PADDY M.R.,
RA SWANSON M.S.;
RN SUBMITTED (JAN-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RN SEQUENCE FROM N.A.
RP RIEGER M., MUELLER-AUER S., SCHAEFER M.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RL -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNP).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL; U05314; G476220; -.
DR EMBL; Z73546; E246917; -.
DR PIR; S48529; S48529.
DR SGD; L0001228; NAB3.
DR PFAM; PF00076; rrm; 1.
KW NAB3-BINDING; NUCLEAR PROTEIN.
FT DOMAIN 68 71 POLY-GLU.
FT DOMAIN 87 93 POLY-GLU.
FT DOMAIN 101 106 POLY-ASP.
FT DOMAIN 108 115 POLY-GLU.
FT DOMAIN 116 127 POLY-ASP.
FT DOMAIN 128 137 POLY-GLU.
FT DOMAIN 603 608 POLY-GLN.
FT DOMAIN 644 648 POLY-PRO.
FT DOMAIN 698 703 POLY-GLN.
FT DOMAIN 723 728 POLY-GLN.
FT DOMAIN 765 768 POLY-PRO.
FT DOMAIN 769 784 POLY-GLN.
SQ SEQUENCE 802 AA; 90438 MW; FB180EDB CRC32;

Query Match 19.98; Score 80; DB 1; Length 802;
Best Local Similarity 32.68; Pred. No. 4.27e-01;
Matches 14; Conservative 10; Mismatches 16; Indels 3; Gaps 3;

Db 622 QGKGSQPPIMNOSYG-RYOTSIPPPPPQOQIQGYGRYQAGP 663
      ||| | :| | :| | :| | :| | :| | :| | :| |
Qy 10 QGY-VQTPF-LSESNRYKISAGSCLPLSTAGPSYVKFDNP 50

RESULT 6
ID POLG_HCV1 STANDARD; PRT; 3011 AA.
AC P26664;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)

```



```

FT DISULFID 25 71 BY SIMILARITY.
FT DISULFID 70 79 BY SIMILARITY.
FT DISULFID 92 108 BY SIMILARITY.
FT DISULFID 107 118 BY SIMILARITY.
FT DISULFID 141 186 BY SIMILARITY.
FT DISULFID 185 194 BY SIMILARITY.
FT DISULFID 216 262 BY SIMILARITY.
FT DISULFID 261 289 BY SIMILARITY.
FT DISULFID 282 296 BY SIMILARITY.
FT DISULFID 295 307 BY SIMILARITY.
FT DISULFID 331 372 BY SIMILARITY.
FT DISULFID 371 380 BY SIMILARITY.
FT DISULFID 403 449 BY SIMILARITY.
FT DISULFID 448 458 BY SIMILARITY.
SQ SEQUENCE 472 AA; 53085 MW; 288949A5 CRC32;

Query Match 19.7%; Score 79; DB 1; Length 472;
Best Local Similarity 30.4%; Pred. No. 6.14e-01;
Matches 14; Conservative 12; Mismatches 17; Indels 3; Gaps 3;

Db 153 DFLYEYSNYGQAPLLVAYTKNY-LSVSGSCCTSANTVCFVK 197
QY 2 EKFMAEFGGYVQTFP-LSESNVRYKISTAGSCPLSTAGP-SYVK 45

RESULT 8
ID VTDB_HUMAN STANDARD; PRT; 474 AA.
AC P02774;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE VITAMIN D-BINDING PROTEIN PRECURSOR (DBP) (GROUP-SPECIFIC COMPONENT)
DE (GC-GLOBULIN) (VDB).
GN GC.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86068030.
RA YANG F., BRUNE J.L., NAYLOR S.L., CUPPLES R.L., NABERHAUS K.H.,
RA BOWMAN B.H.;
RT "Human group-specific component (Gc) is a member of the albumin
RT family";
RL PROC. NATL. ACAD. SCI. U.S.A. 82:7994-7998(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86086396.
RA COOKE N.E., DAVID E.V.;
RT "Serum vitamin D-binding protein is a third member of the albumin and
RT alpha fetoprotein gene family.";
RL J. CLIN. INVEST. 76:2420-2424(1985).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94092730.
RA BRAUN A., KOLER A., MORAWIETZ S., CLEVE H.;
RT "Sequence and organization of the human vitamin D-binding protein
RT gene.";
RL BIOCHIM. BIOPHYS. ACTA 1216:385-394(1993).
RN [4]
RP SEQUENCE OF 17-474.
RX MEDLINE; 86216223.
RA SCHOENTGEN F., METZ-BOUTIQUE M.-H., JOLLES J., CONSTANS J.,
RA JOLLES P.;
RT "Complete amino acid sequence of human vitamin D-binding protein
RT (group-specific component): evidence of a three-fold internal
RT homology as in serum albumin and alpha-fetoprotein.";
RL BIOCHIM. BIOPHYS. ACTA 871:189-198(1986).
RN [5]
RP SEQUENCE OF 17-31 AND 431-441.
RX MEDLINE; 79145448.
RA SVASTI J., KUROSKY A., BENNETT A., BOWMAN B.H.;
RT "Molecular basis for the three major forms of human serum vitamin D

```

```

RT binding protein (group-specific component).";
RL BIOCHEMISTRY 18:1611-1617(1979).
RN [6]
RP SEQUENCE OF 1-19 FROM N.A.
RX MEDLINE; 88005794.
RA YANG F., NABERHAUS K.H., ADRIAN G.S., GARDELLA J.M., BRISSENDEN J.E.,
RA BOWMAN B.H.;
RT "The vitamin D-binding protein gene contains conserved nucleotide
RT sequences that respond to heavy metal, adipocyte and mitotic
RT signals.";
RL GENE 54:285-290(1987).
RN [7]
RP VARIANTS GC2; GC1F AND GC1S.
RX MEDLINE; 92316509.
RA BRAUN A., BICHLMAIER R., CLEVE H.;
RT "Molecular analysis of the gene for the human vitamin-D-binding
RT protein (group-specific component): allelic differences of the common
RT genetic GC types.";
RL HUM. GENET. 89:401-406(1992).
CC -!- FUNCTION: MULTIFUNCTIONAL PROTEIN FOUND IN PLASMA, ASCITIC FLUID,
CC CEREBROSPINAL FLUID, AND URINE AND ON THE SURFACE OF MANY CELL
CC TYPES. IN PLASMA, IT CARRIES THE VITAMIN D STEROLS AND PREVENTS
CC POLYMERIZATION OF ACTIN BY BINDING ITS MONOMERS. DBP ASSOCIATES
CC WITH MEMBRANE-BOUND IMMUNOGLOBULIN ON THE SURFACE OF B-LYMPHOCYTES
CC AND WITH IGG FC RECEPTOR ON THE MEMBRANES OF T-LYMPHOCYTES.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC -!- POLYMORPHISM: OVER 80 VARIANTS OF HUMAN DBP HAVE BEEN IDENTIFIED.
CC THE THREE MOST COMMON ALLELES ARE CALLED GC1F, GC1S, AND GC2. THE
CC SEQUENCE SHOWN IS THAT OF THE GC2 ALLELE.
CC -!- SIMILARITY: BELONGS TO THE ALB/AFP/VDB FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; L10641; G639896; -
EMBL; M11321; G183006; -
EMBL; X03178; G31676; -
EMBL; M12654; G181482; -
EMBL; S67480; G455970; -
EMBL; S67474; G455970; JOINED.
EMBL; S67476; G455970; JOINED.
EMBL; S67478; G455970; JOINED.
EMBL; S67479; G455970; JOINED.
EMBL; S67526; G455970; JOINED.
EMBL; M17156; G463096; -
PIR; A03237; VYHUD.
PIR; A24066; A24066.
PIR; A29096; A29096.
SWISS-2DPAGE; P02774; HUMAN.
MIN; 139200; -
PROSITE; PS00212; ALBUMIN; 1.
PFAM; PF00273; transport_prot; 1.
ALBUMIN; GLYCOPROTEIN; VITAMIN D; TRANSPORT; PLASMA; ACTIN-BINDING;
REPEAT; POLYMORPHISM; SIGNAL.
FT SIGNAL 1 16
FT CHAIN 17 474 VITAMIN D-BINDING PROTEIN.
FT REPEAT 22 199 1.
FT REPEAT 218 385 2.
FT REPEAT 404 474 3.
FT DISULFID 29 75 BY SIMILARITY.
FT DISULFID 74 83 BY SIMILARITY.
FT DISULFID 96 112 BY SIMILARITY.
FT DISULFID 111 122 BY SIMILARITY.
FT DISULFID 145 190 BY SIMILARITY.
FT DISULFID 189 198 BY SIMILARITY.
FT DISULFID 220 266 BY SIMILARITY.
FT DISULFID 265 273 BY SIMILARITY.

```

```
FT DISULFID 286 300 BY SIMILARITY.
FT DISULFID 299 311 BY SIMILARITY.
FT DISULFID 335 376 BY SIMILARITY.
FT DISULFID 375 384 BY SIMILARITY.
FT DISULFID 407 453 BY SIMILARITY.
FT DISULFID 452 462 BY SIMILARITY.
FT CARBOHYD 288 288 POTENTIAL.
FT VARIANT 432 432 D -> E (IN GC15 ALLELE).
FT VARIANT 436 436 K -> T (IN GC1F AND GC1S ALLELES).
FT CONFLICT 168 168 G -> E (IN REF. 2).
FT CONFLICT 327 327 E -> R (IN REF. 2).
SQ SEQUENCE 474 AA; 52963 MW; 56EC123D CRC32;

Query Match 19.7%; Score 79; DB 1; Length 474;
Best Local Similarity 35.7%; Pred. No. 6.14e-01;
Matches 15; Conservative 12; Mismatches 12; Indels 3; Gaps 3;

Db 157 NQFWMEYSTNGQAPLPLVSYTKSY-LSWVGSCC-TSASPT 196
QY 2 EKFAEFGQGYVOTPF-LSESNSVRYKISAGSCPLSTAGPS 42

RESULT 9
ID YQHT_BACSU STANDARD; PRT; 353 AA.
AC P34518;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PUTATIVE PEPTIDASE IN GCVT-SPOIIAA INTERGENIC REGION (EC 3.4.-.-).
GN YQHT.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168 / JH642.
RA KOBAYASHI Y., MIZONO M., MASUDA S., TAKEMARU K., HOSONO S.,
RA SATO T., TAKEUCHI M.;
RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24B.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D84432; D1013236; -
CC DR EMBL; 299116; E1185714; -
CC DR SUBTILIST; BG11708; YQHT.
CC DR PROSITE; PS00491; PROLINE_PEPTIDASE; 1.
CC DR PFAM; PF00557; pep_M24; 1.
CC KW HYPOTHETICAL PROTEIN; HYDROLASE.
CC SQ SEQUENCE 353 AA; 38120 MW; C433A088 CRC32;

Query Match 19.4%; Score 78; DB 1; Length 353;
Best Local Similarity 29.7%; Pred. No. 8.81e-01;
Matches 11; Conservative 14; Mismatches 12; Indels 0; Gaps 0;

Db 3 LEKRNLFQGLGIDMGILTSNTNRYMTGFTGSAGLA 39
QY 1 MEKFAEFGQGYVOTPFELSESNSVRYKISAGSCPLS 37

RESULT 10
ID VTDB_RABIT STANDARD; PRT; 476 AA.
AC P53789;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE VITAMIN D-BINDING PROTEIN PRECURSOR (DBP) (GROUP-SPECIFIC COMPONENT)
```

```
DE (GC-GLOBULIN) (VDB).
GN GC OR DBP.
OS ORYCTOLAGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NEW ZEALAND WHITE; TISSUE-LIVER;
RC MEDLINE; 95218682.
RA OSAWA M., TSUJI T., YUKAWA N., SAITO T., TAKEICHI S.;
RT Cloning and sequence analysis of cDNA encoding rabbit vitamin
RT D-binding protein (gc globulin).;
RL BIOCHEM. MOL. BIOL. INT. 34:1003-1009(1994).
CC -1- FUNCTION: MULTIFUNCTIONAL PROTEIN FOUND IN PLASMA, ASCITIC FLUID,
CC CEREBROSPINAL FLUID, AND URINE AND ON THE SURFACE OF MANY CELL
CC TYPES. IN PLASMA, IT CARRIES THE VITAMIN D STEROLS AND PREVENTS
CC POLYMERIZATION OF ACTIN BY BINDING ITS MONOMERS. DBP ASSOCIATES
CC WITH MEMBRANE-BOUND IMMUNOGLOBULIN ON THE SURFACE OF B-LYMPHOCYTES
CC AND WITH IGG FC RECEPTOR ON THE MEMBRANES OF T-LYMPHOCYTES.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE ALB/AFP/VDB FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D29666; G603499; -
CC DR PROSITE; PS00212; ALBUMIN; 1.
CC DR PFAM; PF00273; transport_prot; 1.
CC KW ALBUMIN; GLYCOPROTEIN; VITAMIN D; TRANSPORT; PLASMA; ACTIN-BINDING;
CC REPEAT; SIGNAL.
CC FT SIGNAL 1 16 BY SIMILARITY.
CC FT CHAIN 17 476 VITAMIN D-BINDING PROTEIN.
CC FT REPEAT 22 199 1.
CC FT REPEAT 218 385 2.
CC FT REPEAT 404 476 3.
CC FT DISULFID 29 75 BY SIMILARITY.
CC FT DISULFID 74 83 BY SIMILARITY.
CC FT DISULFID 96 112 BY SIMILARITY.
CC FT DISULFID 111 122 BY SIMILARITY.
CC FT DISULFID 145 190 BY SIMILARITY.
CC FT DISULFID 189 198 BY SIMILARITY.
CC FT DISULFID 220 266 BY SIMILARITY.
CC FT DISULFID 265 273 BY SIMILARITY.
CC FT DISULFID 286 300 BY SIMILARITY.
CC FT DISULFID 299 311 BY SIMILARITY.
CC FT DISULFID 335 376 BY SIMILARITY.
CC FT DISULFID 375 384 BY SIMILARITY.
CC FT DISULFID 407 453 BY SIMILARITY.
CC FT DISULFID 452 462 BY SIMILARITY.
CC FT CARBOHYD 86 86 POTENTIAL.
CC SQ SEQUENCE 476 AA; 52911 MW; EC2DD7BF CRC32;

Query Match 19.4%; Score 78; DB 1; Length 476;
Best Local Similarity 31.0%; Pred. No. 8.81e-01;
Matches 13; Conservative 15; Mismatches 11; Indels 3; Gaps 3;

Db 157 DKELYESSNYGOAPLPILVSYTKSY-LSWVGTC- TSASPT 196
QY 2 EKFAEFGQGYVOTPF-LSESNSVRYKISAGSCPLSTAGPS 42

RESULT 11
ID POLG_HCVH STANDARD; PRT; 3011 AA.
AC P27958;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
```

```

FT CARBOHYD 576 576 POTENTIAL..
FT CARBOHYD 623 623 POTENTIAL..
FT CARBOHYD 645 645 POTENTIAL..
FT CARBOHYD 2041 2041 POTENTIAL..
FT CARBOHYD 2240 2240 POTENTIAL..
FT CARBOHYD 2364 2364 POTENTIAL..
FT CARBOHYD 2789 2789 POTENTIAL..
SQ SEQUENCE 3011 AA; 327142 MW; 49643481 CRC32;

Query Match      19.4%; Score 78; DB 1; Length 3011;
Best Local Similarity 35.6%; Pred. No. 8.81e-01;
Matches 16; Conservative 12; Mismatches 13; Indels 4; Gaps 4;

Db 2116 IPSPFFETLDGVRHLR-RFAPPCKPILLREEVSRVGLHEYPVGSQ 2159 *
QY 13 VQTPT-FLSENSRVYKISIAAGC-PLSTAGPSY-VXPDNPVGVSQ 54
       :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~
RESULT 12
AC ID NMEL_MOUSE STANDARD; PRT; 1464 AA.
AD P35436;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE DE GLUTAMATE [NMDA] RECEPTOR SUBUNIT EPSILON 1 PRECURSOR (N-METHYL D-ASPARTATE RECEPTOR SUBTYPE 2A) (NR2A) (NMDAR2A).
GN GRIN2A.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
[1]
RN RN SEQUENCE FROM N.A.
RX MEDLINE; 9224361.
RA MEGURO H., MORI H., ARAKI K., KUSHIYA E., KITSUNADA T.,
RA YAMAZAKI M., KOMANISHI T., ARAKAWA M., SAKIMURA K., MISHINA M.;
RT "Functional characterization of a heteromeric NMDA receptor channel expressed from cloned cDNAs.";
RL NATURE 357:70-74(1992).
CC CC -1- FUNCTION: NMDA RECEPTOR SUBTYPE OF GLUTAMATE-GATED ION CHANNELS POSSESSES HIGH CALCIUM PERMEABILITY AND VOLTAGE-DEPENDENT SENSITIVITY TO MAGNESIUM AND IS MEDIATED BY GLYCINE.
CC CC -1- SUBUNIT: HETERODIMER OF AN EPSILON SUBUNIT AND A ZETA SUBUNIT.
CC CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).
-----
CC EMBL; D10217; G220411; ..
CC DR FIR; S29159; S29159.
CC DR MGD; MGI:95820; GRIN2A.
CC DR PFAM; PF00060; lig.Chan. 1.
CC KW RECEPTOR; POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL; TRANSMEMBRANE; CALCIUM; MAGNESIUM.
FT SIGNAL 1 22 POTENTIAL..
FT CHAIN 23 1464 GLUTAMATE [NMDA] RECEPTOR SUBUNIT EPSILON 1.
FT DOMAIN 23 556 EXTRACELLULAR (POTENTIAL)..
FT FT 557 576 1 (POTENTIAL)..
FT TRANSMEM 599 619 2 (POTENTIAL)..
FT TRANSMEM 629 654 3 (POTENTIAL)..
FT TRANSMEM 817 837 4 (POTENTIAL)..
FT SITE 614 614 FUNCTIONAL DETERMINANT OF NMDA RECEPTORS (BY SIMILARITY)..
FT CARBOHYD 75 75 POTENTIAL..
FT CARBOHYD 340 340 POTENTIAL..
FT CARBOHYD 380 380 POTENTIAL..
FT CARBOHYD 443 443 POTENTIAL..

```


Search completed: Wed Sep 1 16:09:53 1999
Job time : 10 secs.

WORLD (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Sep 1 16:10:10 1999; MasPar time 7.61 Seconds
401.470 Million cell updates/sec
Tabular output not generated.

Title: >PCT-US99-13024-2
Description: (1-56) from PCTUS9913024.pep (4 of 12)
Perfect Score: 402
Sequence: 1 MERFMAEFGQGVQTPFLSE.....STAGPSYVKFQDNPVGSQTF 56

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 34.913; Variance 62.861; scale 0.555

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	81	20.1	3011 14	Q03463 GENOME POLYPROTEIN.	1.08e+00
2	80	19.9	275 14	Q68468 POLYPROTEIN (FRAGMENT)	1.50e+00
3	80	19.9	275 14	Q68467 POLYPROTEIN (FRAGMENT)	1.50e+00
4	80	19.9	275 14	Q68469 POLYPROTEIN (FRAGMENT)	1.50e+00
5	80	19.9	357 5	Q01993 VC27A7L.1 PROTEIN	1.50e+00
6	80	19.9	802 3	Q07034 RNA BINDING PROTEIN.	1.50e+00
7	80	19.9	2436 14	Q81756 POLYPROTEIN (FRAGMENT)	1.50e+00
8	79	19.7	542 5	Q17456 SIMILAR TO GLUTAMATE D	2.10e+00
9	79	19.7	860 1	Q59003 860AA LONG HYPOTHETICA	2.10e+00
10	78	19.4	70 2	P75688 FROM BASES 311709 TO 3	2.91e+00
11	78	19.4	3011 14	Q36610 POLYPROTEIN.	2.91e+00
12	78	19.4	3011 14	Q36608 POLYPROTEIN.	2.91e+00
13	78	19.4	3011 14	Q36609 POLYPROTEIN.	2.91e+00
14	78	19.4	3011 14	Q36579 POLYPROTEIN.	2.91e+00
15	77	19.2	418 2	Q56631 LECITHINASE.	4.04e+00
16	77	19.2	470 2	O87325 LECITHINASE.	4.04e+00
17	77	19.2	566 2	O84354 HYPOTHETICAL 63.5 KD P	4.04e+00
18	77	19.2	967 2	Q54123 PEPP.	4.04e+00
19	77	19.2	1464 11	Q63728 N-METHYL-D-ASPARTATE R	4.04e+00
20	77	19.2	1464 11	O08948 N-METHYL-D-ASPARTATE R	4.04e+00

21	77	19.2	1464	4	Q12879	N-METHYL-D-ASPARTATE R	4.04e+00
22	77	19.2	2219	5	Q23388	ZK1057.2 PROTEIN.	4.04e+00
23	76	18.9	535	5	Q91315	SIMILARITY TO HUMAN GT	5.58e+00
24	76	18.9	550	5	O02490	FRUCTOFURANOSIDASE (EC	5.58e+00
25	76	18.9	821	2	O51735	OUTER MEMBRANE PROTEIN	5.58e+00
26	76	18.9	1142	4	O14497	B120.	5.58e+00
27	76	18.9	2660	5	Q19785	F25F2.2 (FRAGMENT).	5.58e+00
28	76	18.9	364	3	O74624	MANNOSE-1-PHOSPHATE GU	7.70e+00
29	74	18.4	178	3	O60074	HYPOTHETICAL 19.4 KD P	1.06e+01
30	74	18.4	921	5	P90770	C34B7.2 PROTEIN	1.06e+01
31	73	18.2	92	14	O72154	L1 PROTEIN (FRAGMENT).	1.45e+01
32	73	18.2	442	4	O43411	HYPOTHETICAL 49.3 KD P	1.45e+01
33	73	18.2	586	9	Q38005	TAIL PROTEIN.	1.45e+01
34	73	18.2	964	2	O53784	PUTATIVE MEMBRANE PROT	1.45e+01
35	73	18.2	1693	14	O81876	COMPLETE GENOME SEQUEN	1.45e+01
36	73	18.2	2301	10	O80784	F13P17.19 PROTEIN.	1.45e+01
37	72	17.9	347	14	O82650	HAEMAGGLUTININ (FRAGME	1.98e+01
38	72	17.9	347	14	O82658	HAEMAGGLUTININ (FRAGME	1.98e+01
39	72	17.9	347	14	O82649	HAEMAGGLUTININ (FRAGME	1.98e+01
40	72	17.9	362	14	Q67383	HAEMAGGLUTININ HAL DOMA	1.98e+01
41	72	17.9	362	14	Q67381	HEMAGGLUTININ (FRAGMEN	1.98e+01
42	72	17.9	379	14	Q67363	HEMAGGLUTININ (FRAGMEN	1.98e+01
43	72	17.9	379	14	Q67366	HEMAGGLUTININ (FRAGMEN	1.98e+01
44	72	17.9	447	14	O89483	POLYPROTEIN (FRAGMENT)	1.98e+01
45	72	17.9	447	14	O89496	POLYPROTEIN (FRAGMENT)	1.98e+01

ALIGNMENTS

RESULT 1
ID Q03463 PRELIMINARY; PRT: 3011 AA.
AC Q03463;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN.
OS HEPATITIS C VIRUS (ISOLATE HC-J1) (HCV).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE;
OC HEPATITIS C-LIKE VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93117120.
RA OKAMOTO H., KANAI N., MISHIRO S.;
RT "Full-length nucleotide sequence of a Japanese hepatitis C virus isolate (HC-J1) with high homology to USA isolates.";
RL NUCLEIC ACIDS RES. 20:6410-6410(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91013116.
RA OKAMOTO H., OKADA S., SUGIYAMA Y., YOTSUMOTO S., TANAKA T.,
RA YOSHIZAWA H., TSUDA F., MIYAKAWA Y., MAYUMI M.;
RT "The 5'-terminal sequence of the hepatitis C virus genome.";
RL JPN. J. EXP. MED. 60:167-177(1990).
DR EMBL; D10749; G221587;
DR PFAM; PF00998; HCV_RDRP; 1.
DR PFAM; PF01001; HCV_NS4b; 1.
DR PFAM; PF01006; HCV_NS4a; 1.
KW POLYPROTEIN.
SQ SEQUENCE 3011 AA; 327114 MW; 399EDBFA CRC32;
Query Match 20.1%; Score 81; DB 14; Length 3011;
Best Local Similarity 40.0%; Pred. No. 1.08e+00;
Matches 18; Conservative 10; Mismatches 13; Indels 4; Gaps 4;
Db 2116 VPSPEFFTELDGVRLLH-RFAPCKPLREVEFSRVLGHDPVGSQ 2159
QY 13 VQTP-FLSESNSRVKISIAAGSC-PLSTAGPSY-VKFDNPVGSQ 54
RESULT 2
ID Q68468 PRELIMINARY; PRT: 275 AA.
AC Q68468;
DT 01-NOV-1996 (TREMREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
 DE POLYPROTEIN (FRAGMENT).
 OS HEPATITIS C VIRUS (HCV).
 OS VIRUSES; SSNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE;
 HEPATITIS C-LIKE VIRUSES.
 [1]
 RC STRAIN-PATIENT #2675;
 RA VIZMANOS J.L.;
 RL THESIS (1996), GENETICS DEPT., UNIVERSITY OF NAVARRA.
 DR EMBL: U37645; G117377; -.
 KW POLYPROTEIN.
 FT NON_TER 1 1
 FT NON_TER 275 275
 SQ SEQUENCE 275 AA; 30504 MW; 6D967905 CRC32;

 Query Match 19.9%; Score 80; DB 14; Length 275;
 Best Local Similarity 37.8%; Pred. No. 1.50e+00;
 Matches 17; Conservative 11; Mismatches 13; Indels 4; Gaps 4;

 Db 140 VSPPEFFTELDGVRHLH-RFAPPCKPLLRREVSFRVGLHEYPVGSQ 183
 QY 13 VQTP-FLUSENSVRKYISAGSC-PLSTAGPSY-VKFDNPNVGSQ 54

 RESULT 3
 ID Q68467 PRELIMINARY; PRT; 275 AA.
 AC Q68467;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DE POLYPROTEIN (FRAGMENT).
 OS HEPATITIS C VIRUS (HCV).
 OS VIRUSES; SSNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE;
 HEPATITIS C-LIKE VIRUSES.
 [1]
 RC STRAIN-PATIENT #2675;
 RA VIZMANOS J.L.;
 RL THESIS (1996), GENETICS DEPT., UNIVERSITY OF NAVARRA.
 DR EMBL: U37644; G117377; -.
 KW POLYPROTEIN.
 FT NON_TER 1 1
 FT NON_TER 275 275
 SQ SEQUENCE 275 AA; 30423 MW; 6DE9E548 CRC32;

 Query Match 19.9%; Score 80; DB 14; Length 275;
 Best Local Similarity 37.8%; Pred. No. 1.50e+00;
 Matches 17; Conservative 11; Mismatches 13; Indels 4; Gaps 4;

 Db 140 VSPPEFFTELDGVRHLH-RFAPPCKPLLRREVSFRVGLHEYPVGSQ 183
 QY 13 VQTP-FLUSENSVRKYISAGSC-PLSTAGPSY-VKFDNPNVGSQ 54

 RESULT 4
 ID Q68469 PRELIMINARY; PRT; 275 AA.
 AC Q68469;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DE POLYPROTEIN (FRAGMENT).
 OS HEPATITIS C VIRUS (HCV).
 OS VIRUSES; SSNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE;
 HEPATITIS C-LIKE VIRUSES.
 [1]
 RC STRAIN-PATIENT #2675;
 RA VIZMANOS J.L.;
 RL THESIS (1996), GENETICS DEPT., UNIVERSITY OF NAVARRA.
 DR EMBL: U37646; G117379; -.
 KW POLYPROTEIN.

FT NON_TER 1 1
 FT NON_TER 275 275
 SQ SEQUENCE 275 AA; 30495 MW; 9BE1EE8B CRC32;

 Query Match 19.9%; Score 80; DB 14; Length 275;
 Best Local Similarity 37.8%; Pred. No. 1.50e+00;
 Matches 17; Conservative 11; Mismatches 13; Indels 4; Gaps 4;

 Db 140 VSPPEFFTELDGVRHLH-RFAPPCKPLLRREVSFRVGLHEYPVGSQ 183
 QY 13 VQTP-FLUSENSVRKYISAGSC-PLSTAGPSY-VKFDNPNVGSQ 54

 RESULT 5
 ID O01993 PRELIMINARY; PRT; 357 AA.
 AC O01993;
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE VC27A7L.1 PROTEIN.
 GN VC27A7L.1
 OS CAENORHABDITIS ELEGANS.
 OC RHABDITINA; RHABDITOIDEA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
 OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 [1]
 RN SEQUENCE FROM N.A.
 RA WHITE S.;
 RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE: 94150718.
 RX WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BOWFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL NATURE 368:32-38(1994).
 DR EMBL: Z95621; E1349927; -.
 SQ SEQUENCE 357 AA; 41307 MW; D2FB4683 CRC32;

 Query Match 19.9%; Score 80; DB 5; Length 357;
 Best Local Similarity 45.8%; Pred. No. 1.50e+00;
 Matches 11; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

 Db 211 RSYLETPFYRDYNSDVYKVPILGS 234
 QY 10 QGYVQTPFLSESNVRKYISAGS 33

 RESULT 6
 ID Q07034 PRELIMINARY; PRT; 802 AA.
 AC Q07034;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE RNA BINDING PROTEIN.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
 OC SACCHAROMYCETACEAE; SACCHAROMYCES.
 [1]
 RN SEQUENCE FROM N.A.
 RA SUGIMOTO K., MATSUMOTO K., KORNBERG R.D., REED S.I., WITTENBERG C.;
 RT "HMDI of the yeast Saccharomyces cerevisiae encodes a putative
 RT RNA/single strand DNA-binding protein, whose overexpression reduces
 RT the expression of a g1 cyclin, CLN3."
 RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.

[2]
RN SEQUENCE FROM N.A.
RP STRAIN-BRISTOL N2;
RC DU Z., GATTUNG S.;
RA SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
[3]
RN SEQUENCE FROM N.A.
RP STRAIN-BRISTOL N2;
RC WATERSTON R.;
RA SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U50312; G1226312; -;
DR PFAM; PFO0282; pyridoxal dec; 1.
SQ SEQUENCE 542 AA; 61133 MW; A7C148AB CRC32;

Query Match 19.7%; Score 79; DB 5; Length 542;
Best Local Similarity 30.3%; Pred.No. 2.10e+00;
Matches 10; Conservative 11; Mismatches 10; Indels 2; Gaps 2

Db 309 GFLP-FLEE-DEIRYDFRVGVSISADSHKY 339
|::|||::|::|::|:
QY 11 GYVTFPLESNVYKISAGSCPLSTAGPSY 43
|::|||::|::|::|:

RESULT 9
ID OS9003 PRELIMINARY; PRT; 860 AA.
AC OS9003;
DT 01-AUG-1998 (TREMBREL. 07, CREATED)
DT 01-AUG-1998 (TREMBREL. 07, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBREL. 09, LAST ANNOTATION UPDATE)
DE 860AA LONG HYPOTHETICAL PROTEIN.
GN PH1302.
OS PYROCOCUS HORIKOSHII.
OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCCUS.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-OT3;
RC MEDLINE; 98344137.
RA KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAICAWA Y., HINO Y.,
RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOVAMA A., NAGAI Y.,
RA SAKAI M., OGURA K., OTUKA K., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT *Complete Sequence and Gene Organization of the Genome of a
Hyperthermophilic Archaeobacterium, Pyrococcus horikoshii OT3.*;
RL DNA RES. 5:55-76(1998).
DR EMBL; AP000006; D1031349; -;
SQ SEQUENCE 860 AA; 95455 MW; B336F868 CRC32;

Query Match 19.7%; Score 79; DB 1; Length 860;
Best Local Similarity 28.0%; Pred.No. 2.10e+00;
Matches 7; Conservative 10; Mismatches 8; Indels 0; Gaps 0

Db 1 MSKFLTDFSNAFIVAFKINSYEAR 25
|::|::|::|:
QY 1 MEKFMAEFGGYVQTPLSESVNR 25
|::|::|::|:

RESULT 10
ID P75688 PRELIMINARY; PRT; 70 AA.
AC P75688;
DT 01-FEB-1997 (TREMBREL. 02, CREATED)
DT 01-FEB-1997 (TREMBREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE FROM BASES 311709 TO 323910
DE (SECTION 27 OF 400) OF THE COMPLETE GENOME (SECTION 27 OF 400).
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-K-12;
RX MEDLINE; 97426617.

Search completed: Wed Sep 1 16:10:34 1999
Job time : 24 secs.

Result No.	Score	Query		Length	DB	ID	Description	Pred. No
		Match	%					
1	300	54.4	65.9	19	R97371	R73368	Phage T4-ORF7-gene pr	1.22e-19
2	96	17.4	1693	35	R73368	R73368	Hepatitis E virus hol	1.66e-00
3	87	15.8	623	35	R57573	R57573	Amino acid sequence o	9.05e-00
4	86	15.6	1693	10	R51264	R51264	HEV strain protein en	1.09e-01
5	86	15.6	1693	36	W80196	W80196	Protein encoded by OR	1.09e-01
6	86	15.6	1693	34	R91813	R91813	Hepatitis E virus str	1.09e-01
7	86	15.6	1693	37	R71209	R71209	Protein encoded by OR	1.09e-01
8	86	15.6	1693	37	W81519	W81519	Hepatitis E virus (HE	1.09e-01
9	85	15.4	633	35	R57574	R57574	Amino acid sequence o	1.31e-01
10	84	15.2	1693	3	R14618	R14618	Protein encoded by OR	1.57e-01
11	82	14.9	633	35	R57575	R57575	Amino acid sequence o	2.27e-01
12	81	14.7	3011	12	R66995	R66995	Hepatitis C virus gen	2.72e-01
13	81	14.7	3011	8	R40120	R40120	HCV genomic amino aci	2.72e-01
14	80	14.5	170	1	P92033	P92033	Sequence encoded in t	3.26e-01
15	80	14.5	170	1	P90150	P90150	Sequence of hepatitis	3.26e-01
16	80	14.5	411	16	R90934	R90934	HCV NS5 domain antige	3.26e-01

05-JUL-1990; US-505888.
07-JUN-1995; US-475807.
FA (GENE-) GENELABS TECHNOLOGIES INC.
PI Bradley DW, Fry KE, Krawczynski KZ, Reyes GR, Tam A,
PI Yarbough PO;
DR WPI: 98-582599/49.
DR N-PSDB: V66321.
PT Hepatitis E virus proteins - useful for diagnosis or vaccine
production the virus
PS Claim 22; Columns 57-66; 47pp; English.
CC W80196-98 are encoded by the genome of the Burma strain of
CC enterically transmitted non-A non-B hepatitis virus (ET-NANB)
CC (hepatitis E virus (HEV)). The specification describes an isolated
CC protein which is specifically immunoreactive with antibodies present
CC in individuals infected with HEV and encoded by a sequence contained
CC in an open reading frame (ORF) of an HEV genome. The genome has a
CC sequence that is more than 70% identical to the ORF1 sequence from
CC Burma HEV isolate. The protein is used as a vaccine and a diagnostic
CC probe for ET-NANB.
CC Query Match 15.6%; Score 86; DB 36; Length 1693;
Best Local Similarity 25.4%; Pred. No. 1.09e+01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;
SQ Sequence 1693 AA;

Db 572 frtsfdgavltnqperhnlsfdaqstmaagpfsltyasaaglevryvaagldhrav 631
Qy 8 FCQGYYVTPLSESNRYKISAGSCPLSTAGSPYVKFQDNPVGSGT-F-SAGLHLR-V 64
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 632 fapgvsp 638
Qy 65 FDPSTGA 71
| : : |

RESULT 6
ID R91813 standard; Protein; 1693 AA.
AC R91813;
DT 26-NOV-1996 (first entry)
DE Hepatitis E virus strain SAR-55 ORF-1.
KW Hepatitis E virus; HEV; SAR-55 strain; enteric transmission;
KW structural region; antigen; detection; antibody; vaccine;
KW immunisation; infection.
OS Hepatitis E virus.
FH Key Location/Qualifiers
FT misc_difference 1238 /note= "corresponding codon CAG"
FT misc_difference 1455..1693 /note= "10 bp nucleic acid sequence TGGTNTTYGA has to be inserted between nucleotides 4390..4391 of T27394 before these amino acid residues can be decoded"
FT FT
FT FT
FT FT
PN W09610580-A2.
PD 11-APR-1996.
PF 03-OCT-1995; U13102.
PR 03-OCT-1994; US-316765.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Emerson SU, Purcell RH, Tsarev SA;
DR WPI: 96-209320/21.
DR N-PSDB: T27394.
PT Isolated and purified hepatitis E virus strain SAR-55 DNA - encodes antigenic protein useful in diagnosis, prophylaxis and treatment of hepatitis E virus infection
PS Disclosure; pages 9-13; 121pp; English.
CC The present sequence is the protein prod. of ORF-1 from the hepatitis E virus (HEV) strain SAR-55, which was implicated in an enterically transmitted non-A, non-B hepatitis in Pakistan. The protein encoded by the structural region of the virus (i.e. ORF-2), which is capable of forming HEV like particles, is useful for the detection of HEV antibodies (pref. IgG or IgM) in blood, plasma, sera, cerebrospinal fluid, tissue, urine or pleural fluid. The protein, and anti-HEV antibodies generated using the protein, can also be used in vaccines for immunising an animal against HEV infection. The protein is identified as a band of greater than

05-KD following SDS-PAGE of cell lysates of insect cells infected with a HEV ORF-2 contg. baculovirus, i.e. the claimed recombinant expression vectors pPIC9-1779, -1780 and -1781.
SQ Sequence 1693 AA;

Query Match 15.6%; Score 86; DB 19; Length 1693;
Best Local Similarity 25.4%; Pred. No. 1.09e+01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;
SQ Sequence 1693 AA;

Db 572 frtsfdgavltnqperhnlsfdaqstmaagpfsltyasaaglevryvaagldhrav 631
Qy 8 FCQGYYVTPLSESNRYKISAGSCPLSTAGSPYVKFQDNPVGSGT-F-SAGLHLR-V 64
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 632 fapgvsp 638
Qy 65 FDPSTGA 71
| : : |

RESULT 7
ID W71209 standard; Protein; 1693 AA.
AC W71209;
DT 30-OCT-1998 (first entry)
DE Protein encoded by ORF 1 of the Burmese isolate of ET-NANB.
KW Enterically transmitted nonA/nonB hepatitis virus; identification; HEV; ET-NANB; detection; vaccine.
OS Hepatitis virus.
FH Key Location/Qualifiers
FT Misc_difference 154 /note= "not specified"
FT Misc_difference 1514 /note= "not specified"
FT Misc_difference 1552 /note= "not specified"
FT FT
PN US5789559-A.
PD 04-AUG-1998.
PF 25-JUN-1994; 279823.
PR 05-APR-1991; US-681078.
PR 17-JUN-1988; US-208997.
PR 11-APR-1989; US-336672.
PR 19-JUN-1989; US-367486.
PR 13-OCT-1989; US-420921.
PR 05-APR-1990; US-505888.
PR 25-JUL-1994; US-279823.
PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Bradley DW, Fry KE, Krawczynski KZ, Reyes GR, Tam A,
PI Yarbough PO;
DR WPI: 98-446186/38.
DR N-PSDB: V54729.
PT Hepatitis E virus DNA - useful for e.g. virus detection and viral protein production
PS Disclosure; Columns 55-64; 45pp; English.
CC W71209-11 represent the proteins encoded by the open reading frames (ORFs) of the DNA sequence of the Burmese isolate of an enterically transmitted nonA/nonB viral hepatitis agent (ET-NANB). The nucleic acid sequence may be used for identifying and sequencing the entire viral agent (also referred to as HEV), detecting ET-NANB in infected samples, e.g. by specific amplification of virus-derived DNA sequences and for producing recombinant viral proteins for use in vaccines.
CC Query Match 15.6%; Score 86; DB 34; Length 1693;
Best Local Similarity 25.4%; Pred. No. 1.09e+01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;
SQ Sequence 1693 AA;

Db 572 frtsfdgavltnqperhnlsfdaqstmaagpfsltyasaaglevryvaagldhrav 631
Qy 8 FCQGYYVTPLSESNRYKISAGSCPLSTAGSPYVKFQDNPVGSGT-F-SAGLHLR-V 64
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 632 fapgvsp 638
Qy 65 FDPSTGA 71
| : : |

Query Match 15.6%; Score 86; DB 34; Length 1693;
Best Local Similarity 25.4%; Pred. No. 1.09e+01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;
SQ Sequence 1693 AA;

Db 572 frtsfdgavltnqperhnlsfdaqstmaagpfsltyasaaglevryvaagldhrav 631
Qy 8 FCQGYYVTPLSESNRYKISAGSCPLSTAGSPYVKFQDNPVGSGT-F-SAGLHLR-V 64
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 632 fapgvsp 638
Qy 65 FDPSTGA 71
| : : |

[illegible]


```

Db      63  vpspefteldgvrh-rfapckpllreevsfrvlgheypvgsq 106
      - : : : : : : : : : : : : : : : : : : : : : : : :
QY     13  VQTP-FLSESNSVRYKISITAGSC-PLSTAGPSY-VKFDQNPVGSQ  54

```


Search completed: wed Sep 1 16:15:06 1999
Job time : 37 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

mfsrch_pp  protein - protein database search, using Smith-Waterman algorithm
Run on:      Wed Sep 1 16:13:51 1999;  MasPar time 6.25 seconds
              500.410 Million cell updates/sec
Tabular output not generated.

```

```
>PCT-US99-13024-2
Title:
Description:
Perfect Score:
1 (1-78) from PCTUS9913024.pep (5 of 12)
551
Sequence: 1 MEKFMNAEFGQGYVQTFPFLSE.....GLHLRFDPSTGALVDKSY 78
```

Scoring table: PAM 150
Gap 11

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60

Statistics: Mean 37.846; Variance 68.015; scale 0.556

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB	ID	Description	Pred. No.
	Score	Match	Length				
1	92	16.7	299	2	S61248	hypothetical protein	7.69e-02
2	89	16.2	566	2	G17525	probable OMP [leader	2.14e-01
3	86	15.6	1693	1	MNWHE	genome polyprotein -	5.84e-01
4	85	15.4	331	2	A61046	ecdysone-induced memb	8.12e-01
5	85	15.4	869	2	C56617	cfac protein precursor	8.12e-01
6	84	15.2	622	2	S17402	parasporal crystal pr	1.13e+00
7	83	15.1	281	2	S38913	hypothetical protein	1.56e+00
8	83	15.1	611	2	D70928	hypothetical protein	1.56e+00
9	82	14.9	81	2	I38375	tyrosine kinase - hum	2.15e+00
10	82	14.9	449	2	S55092	hypothetical protein	2.15e+00
11	82	14.9	527	2	I84483	tyrosine kinase - hum	2.15e+00
12	81	14.7	216	2	G70447	flagellar L-ring prot	2.96e+00
13	81	14.7	527	2	A55631	protein-tyrosine kina	2.96e+00
14	81	14.7	527	2	I49133	Txk - mouse	2.96e+00
15	81	14.7	3011	2	S40770	polyprotein precursor	2.96e+00
16	80	14.5	171	2	B38162	hypothetical protein	4.05e+00
17	80	14.5	572	2	S14200	GRESAG protein - Tryp	4.05e+00
18	80	14.5	802	2	A48529	NAB3 protein - yeast	4.05e+00
19	80	14.5	954	2	S46105	glucan 1,4-alpha-gluc	4.05e+00
20	80	14.5	1659	2	JC4956	vitellogenin precursor	4.05e+00
21	80	14.5	3011	1	GNWVC3	genome polyprotein -	4.05e+00
22	79	14.3	370	2	S69178	hypothetical protein	5.55e+00
23	79	14.3	472	1	A35327	vitamin D-bindingpro	5.55e+00

24	79	14.3	474	1	VYHUD	vitamin D-binding pro	5.55e+00
25	79	14.3	474	2	S39787	vitamin D-binding pro	5.55e+00
26	79	14.3	633	2	D32053	parapsoral crystal pr	5.55e+00
27	79	14.3	860	2	F71000	hypothetical protein	5.55e+00
28	79	14.3	935	2	S57080	hypothetical protein	5.55e+00
29	78	14.2	70	2	E64757	probable membrane pr	7.57e+00
30	78	14.2	353	2	C69660	Xaa-Pro dipeptidase h	7.57e+00
31	78	14.2	590	2	A29913	parapsoral crystal pr	7.57e+00
32	78	14.2	613	2	T00758	ethylene response sen	7.57e+00
33	78	14.2	633	2	C32053	parapsoral crystal pr	7.57e+00
34	78	14.2	687	2	S56939	probable membrane pr	7.57e+00
35	78	14.2	3011	1	GNWVCH	genome polyprotein -	7.57e+00
36	77	14.0	122	2	S40198	ribosomal protein L14	1.03e-01
37	77	14.0	497	2	I46692	cholesteryl ester tra	1.03e-01
38	77	14.0	967	2	S58360	pepB protein - Staphy	1.03e-01
39	77	14.0	1363	1	VG1HNM	E2 glycoprotein precu	1.03e-01
40	77	14.0	1464	2	S29159	glutamate receptor, N	1.03e-01
41	77	14.0	1464	2	S47555	N-methyl-D-aspartate	1.03e-01
42	77	14.0	1464	2	A43274	N-methyl D-aspartate	1.03e-01
43	76	13.8	338	2	A45247	tcpF protein precursor	1.39e-01
44	76	13.8	821	2	B70199	outer membrane protei	1.39e-01
45	76	13.8	1142	2	T00022	Bl20 protein - human	1.39e-01

ALIGNMENTS

RESULT	1
ENTRY	S61248 #type complete
TITLE	hypothetical protein - bovine herpesvirus 1
ORGANISM	#formal_name bovine herpesvirus 1
DATE	18-Sep-1997 #sequence_revision 18-Sep-1997 #text_change 21-Aug-1998
ACCESSIONS	S61248
REFERENCE	S61233
AUTHORS	Vlcek, C.; Benes, V.; Lu, Z.; Kutish, G.F.; Paces, V.; Rock, D.; Letchworth, G.J.; Schwytzer, M.
SUBMISSION	submitted to the EMBL Data Library, January 1995
DESCRIPTION	Nucleotide sequence analysis of a 30-kb region of the bovine herpesvirus 1 genome which exhibits a colinear gene arrangement with the UL21 to UL4 genes of herpes simplex virus.
ACCESION	S61248
STATUS	Preliminary
MOLECULE_TYPE	DNA
RESIDUES	1-299 #label VLC
CROSS_REFERENCES	EMBL:D48053; NID:g971311; PID:g971327
CLASSIFICATION	#superfamily varicella-zoster virus gene 53 protein
LENGTH	299 #molecular_weight 32379 #checksum 1192
QUERY MATCH	Query Match 16.7%; Score 92; DB 2; Length 299;
BEST LOCAL SIMILARITY	Best Local Similarity 47.2%; Pred. No. 7,69e+02;
MATCHES	Conservative 9; Mismatches 6; Indels 4; Gaps 4; Matches 17;
DBD	45 PRFVCEVEIRPAGPTFTSSITHLRV-EPSTGALL 79 : :: : : :: :
OY	41 PSYV-KFDONPVGSQTF-SAGL-HLRVPDPSTGALV 73 : :: : : :: :
RESULT	2
ENTRY	G71525 #type complete
TITLE	probable OMP [leader (19) peptide] - Chlamydia trachomatis (serotype D, strain UW3/Cx)
ORGANISM	#formal_name Chlamydia trachomatis
DATE	13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 21-Nov-1998
ACCESSIONS	G71525
REFERENCE	A71570
AUTHORS	Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W. Science (1998) 282:754-759
JOURNAL	Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.
TITLE	

```
#cross-references MUID:99000809
#accession G71525
##status preliminary
##molecule_type DNA
##residues 1-566 ##label ARN
##cross-references GB:AE001308; GB:AE001273; NID:93328766; PID:g3328771
##experimental_source serotype D, strain UW-3/Cx
GENETICS
#gene
#accession CT350
#length 566 #molecular-weight 63507 #checksum 4960
SUMMARY
Query Match 16.2%; Score 89; DB 2; Length 566;
Best Local Similarity 26.2%; Pred. No. 2.14e-01;
Matches 17; Conservative 25; Mismatches 18; Indels 5; Gaps 5;
Db 63 IAEYLOQSFLESDTVIR-KSAIIGA-GLSGSEA-LELLSEAIETQDLYEQI-L-ILNA 117
QY 8 FCGYVQTFPLESSESVRYKISAGSCPLSTAGPSYVKFQDNPVGSQTFSAGLHLRVFDP 67
Db 118 ATSQL 122
QY 68 STGAL 72
RESULT 3
ENTRY MNWHE #type complete
TITLE genome polyprotein - hepatitis E virus (strain Burma)
CONTAINS RNA-directed RNA polymerase (EC 2.7.7.48)
ORGANISM #formal_name hepatitis E virus
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
29-May-1998
ACCESSIONS A40778; A48547
REFERENCE A40778
#authors Tam, A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes, G.R.
#journal Virology (1991) 185:120-131
#title Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome.
#cross-references MUID:92024067
#accession A40778
##molecule_type genomic RNA
##residues 1-1693 ##label TAM
##cross-references GB:M73218; NID:9330023; PID:g3330024
REFERENCE A48547
#authors Fry, K.E.; Tam, A.W.; Smith, M.M.; Kim, J.P.; Luk, K.C.; Young, L.M.; Piatk, M.; Feldman, R.A.; Yun, K.Y.; Purdy, M.A.; McCaustland, K.A.; Bradley, D.W.; Reyes, G.R.
#journal Virus Genes (1992) 6:173-185
#title Hepatitis E virus (HEV): strain variation in the nonstructural gene region encoding consensus motifs for an RNA-dependent RNA polymerase and an ATP/GTP binding site.
#cross-references MUID:92271462
#accession A48547
##molecule_type genomic RNA
##residues 967-1693 ##label FRY
##cross-references GB:M32400; NID:9330021; PID:g3330022
##note sequence extracted from NCBI backbone (NCBIN:104572, NCBI:P:104573)
CLASSIFICATION #superfamily hepatitis E virus nonstructural protein
KEYWORDS ATP; nonstructural protein; nucleotidyltransferase
SUMMARY #length 1693 #molecular-weight 185191 #checksum 6520
Query Match 15.6%; Score 86; DB 1; Length 1693;
Best Local Similarity 25.4%; Pred. No. 5.84e-01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;
Db 572 FRTSVDGAVLTNGPERNLNLSFDASQSTMAAGPPLSLTYAASAGLEVYVVAAGLDHRAV 631
QY 8 FCGYVQTFPLESSESVRYKISAGSCPLSTAGPSYVKFQDNPVGSQT-F-SAGLHLR-V 64
Db 632 FAPGVSP 638
QY 65 FDPSTGA 71
RESULT 4
ENTRY A61046 #type complete
TITLE ecdysone-induced membrane protein IMP-E3 - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 16-Feb-1997
ACCESSIONS A61046
REFERENCE A61046
#authors Moore, J.T.; Fristrom, D.; Hammonds, A.S.; Fristrom, J.W.
#journal Dev. Genet. (1990) 11:299-309
#title Characterization of IMP-E3, a gene active during imaginal disc morphogenesis in Drosophila melanogaster.
#accession A61046
##status preliminary
##molecule_type mRNA
##residues 1-331 ##label MOO
GENETICS
#gene FlyBase:ImpE3
##cross-references FlyBase:FBgn0001255
KEYWORDS membrane protein
SUMMARY #length 331 #molecular-weight 36583 #checksum 8221
Query Match 15.4%; Score 85; DB 2; Length 331;
Best Local Similarity 31.9%; Pred. No. 8.12e-01;
Matches 15; Conservative 12; Mismatches 18; Indels 2; Gaps 2;
Db 190 LNFRLYLDNDYGRAFGSESAMDRTASIAAGKRVPTKP-YVDF 235
QY 1 MEKFNAEFGQGVQVTFPLESSESVRYKI-SIAGSCPLSTAGPSYVKF 46
RESULT 5
ENTRY C56617 #type complete
TITLE cfac protein precursor - Escherichia coli plasmid NTP113
ORGANISM #formal_name Escherichia coli
DATE 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 20-Mar-1998
ACCESSIONS C56617
REFERENCE A56617
#authors Jordi, B.J.; Willshaw, G.A.; van der Zeijst, B.A.; Gaastra, W.
#journal DNA Seq. (1992) 2:257-263
#title The complete nucleotide sequence of region 1 of the CFA/I fimbrial operon of human enterotoxigenic Escherichia coli.
#cross-references MUID:92329981
#accession C56617
##status preliminary
##molecule_type DNA
##residues 1-869 ##label JOR
##cross-references GB:M55861; NID:g145507; PID:g145510
##experimental_source enterotoxigenic strain, CFA/I-St plasmid NTP113
##note sequence extracted from NCBI backbone (NCBIN:108960, NCBI:P:108971)
GENETICS
#gene cfac
#genome plasmid
SUMMARY #length 869 #molecular-weight 97830 #checksum 9755
Query Match 15.4%; Score 85; DB 2; Length 869;
Best Local Similarity 32.4%; Pred. No. 8.12e-01;
Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;
Db 158 AFIQSQTNLSDSGYKRLISGNSALGTDTSY 191
QY 11 GYVQTFPLESSESVRYK-ISIAGSCPLSTAGPSY 43
RESULT 6
ENTRY SI7402 #type complete
TITLE parasporal crystal protein cryIIC - Bacillus thuringiensis
```

```

plasmid
ALTERNATE_NAMES  delta-endotoxin
ORGANISM          #formal_name Bacillus thuringiensis
DATE             22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
                09-Sep-1997
ACCESSIONS       SI17402
REFERENCE        SI17400
#authors        Wu, D.; Cao, X.L.; Bai, Y.Y.; Aronson, A.I.
#journal        FEMS Microbiol. Lett. (1991) 81:31-36
#title          Sequence of an operon containing a novel delta-endotoxin gene
                from Bacillus thuringiensis.
#accession      SI17402
##status        preliminary
##molecule_type DNA
##residues      1-622 ##label WUD
##cross-references EMBL:X57252; NID:g40283; PID:g40286
GENETICS
#genome         plasmid
KEYWORDS         delta-endotoxin
SUMMARY          #length 622 #molecular-weight 69729 #checksum 8205
Query Match     15.2%; Score 84; DB 2; Length 622;
Best Local Similarity 18.2%; Pred. No. 1.13e+00;
Matches 12; Conservative 26; Mismatches 26; Indels 2; Gaps 2;
Db 501 FISEYKNGDLSRELNPRTARYTLRNGNSYNYLRVSSIGSTIRVTINGRYTANV 560
QY 12 VYQTPFLSESNVRYKISAGS-CPLSTAGPSYVKF-QDNPGVSGTFSAGLHLRVDPST 69
Db 561 WTTTNN 566
QY 70 GALVDS 75

RESULT 7
ENTRY  S38913 #type complete
TITLE  hypothetical protein 1 - phage phi-C31
ORGANISM #formal_name phage phi-C31
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
09-Sep-1997
ACCESSIONS S38913
REFERENCE S38912
#authors Hartley, N.M.; Murphy, G.O.; Bruton, C.J.; Chater, K.F.
#submission submitted to the EMBL Data Library, November 1993
#accession S38913
##status preliminary
##molecule_type DNA
##residues 1-281 ##label HAR
##cross-references EMBL:X76288; NID:g432610; PID:g579071
GENETICS
#start_codon GTG
SUMMARY #length 281 #molecular-weight 31680 #checksum 2790
Query Match 15.1%; Score 83; DB 2; Length 281;
Best Local Similarity 24.1%; Pred. No. 1.56e+00;
Matches 14; Conservative 17; Mismatches 24; Indels 3; Gaps 3;
Db 194 AYMNADFIIDGNREPMPEFDGAALVHVTDTWA-FKPVTGPDVFAQFLHLRQTFD 250
QY 11 GYVQTPFLSESNVRYKIS-ITAGSCPLSTAGPSYVKFQDNPGVSGTFSAGLHLR-VFD 66

RESULT 8
ENTRY  D70928 #type complete
TITLE  hypothetical protein Rv2913c - Mycobacterium tuberculosis
                (strain H37Rv)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS D70928
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
                C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry

```

```

III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
Nature (1998) 393:537-544
Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
#accession D70928
##status preliminary; nucleic acid sequence not shown;
                translation not shown
##molecule_type DNA
##residues 1-611 ##label COL
##cross-references GB:274024; GB:AL123456; NID:g3250700; PID:e1301028;
                PID:g3250707
##experimental_source strain H37Rv
GENETICS
#gene Rv2913c
SUMMARY #length 611 #molecular-weight 67204 #checksum 9971
Query Match 15.1%; Score 83; DB 2; Length 611;
Best Local Similarity 43.5%; Pred. No. 1.56e+00;
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
Db 319 VRFQHLVPFFELYSDGIDLVPFE 341
QY 44 VKFDQNPVGSQTFSGAGLHLRVFD 66

RESULT 9
ENTRY  I38375 #type fragment
TITLE  tyrosine kinase - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 29-May-1998 #sequence_revision 29-May-1998 #text_change
10-Jul-1998
ACCESSIONS I38375
REFERENCE I38372
#authors Haile, R.N.; Ohta, Y.; Lewis, J.E.; Fu, S.M.; Kroisel, P.;
                Litman, G.W.
#journal Hum. Mol. Genet. (1994) 3:897-901
#title TXK, a novel human tyrosine kinase expressed in T cells
                shares sequence identity with Tec family kinases and maps
                to 4p12.
#cross-references MUID:95038742
#accession I38375
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type DNA
##residues 1-81 ##label RES
##cross-references EMBL:U07794; NID:g508219; PID:g508224
GENETICS
#gene GDB:TXK
#cross-references GDB:377329; OMIM:600058
#map_position 4p12-4p12
#introns 25/1
CLASSIFICATION #superfamily protein-tyrosine kinase tec; pleckstrin repeat
                homology; protein kinase homology; SH2 homology; SH3
                homology
FEATURE
#domain domain protein kinase homology (fragment) #label KIN
SUMMARY #length 81 #checksum 7751
Query Match 14.9%; Score 82; DB 2; Length 81;
Best Local Similarity 44.4%; Pred. No. 2.15e+00;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
Db 5 LRYPVGLMGSLCPATAGSYEKWEIDP 31
QY 24 VRYKISAGSCPLSTAGPSYVKFQDN 50

```

```
RESULT 10
ENTRY
TITLE      S55092      #type complete
            hypothetical protein YMR210w - yeast (Saccharomyces
            cerevisiae)
ALTERNATE_NAMES
ORGANISM    #formal_name Saccharomyces cerevisiae
DATE        08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change
            04-Sep-1998
ACCESSIONS  S55092
REFERENCE    S55089      #authors
            Dedman, K.; Brown, D.; Bowman, S.
            #submission submitted to the EMBL Data Library, June 1995
            #accession S55092
            ##molecule_type DNA
            ##residues 1-449 #label DED
            ##cross-references EMBL:249809; NID:g954459; PID:g954462; MIPS:YMR210w
            ##experimental_source strain AB972
GENETICS
            #map_position 13R
            #superfamily conserved hypothetical protein YBR177c
            #length 449 #molecular-weight 51437 #checksum 3020
SUMMARY
Query Match 14.9%; Score 82; DB 2; Length 449;
Best Local Similarity 27.1%; Pred. No. 2.15e+00;
Matches 16; Conservative 18; Mismatches 22; Indels 3; Gaps 3;
Db 227 YMGFSIGASIMTYLGEESDRTKIECAISVSNPFDLYNSAYF-INSTPMGSRFYSPAL 284
QY 4 FMAEFGQG-YVQTFPLES-SNSVRKYISAGSCPLSTAGPSYVKFQDNPVGSGTFSAGL 60

RESULT 11
ENTRY
TITLE      I84483      #type complete
            tyrosine kinase - human
ORGANISM    #formal_name Homo sapiens #common_name man
DATE        29-May-1998 #sequence_revision 29-May-1998 #text_change
            12-Feb-1999
ACCESSIONS  I84483
REFERENCE    I38372
            #authors
            Haire, R.N.; Ohta, Y.; Lewis, J.E.; Fu, S.M.; Kroisel, P.;
            Litman, G.W.
            #journal Hum. Mol. Genet. (1994) 3:897-901
            #title TXK, a novel human tyrosine kinase expressed in T cells
            #title shares sequence identity with Tec family kinases and maps
            #title to 4p12
            #cross-references MUID:95038742
            #accession I84483
            ##status preliminary; translated from GB/EMBL/DBDJ
            ##molecule_type mRNA
            ##residues 1-527 #label RES
            ##cross-references GB:L27071; NID:g951045; PID:g684986
GENETICS
            #gene GDB:TXK
            ##cross-references GDB:377329; OMIM:600058
            #map_position 4p12-4p12
            #superfamily protein-tyrosine kinase tec; pleckstrin repeat
            #homology; protein kinase homology; SH2 homology; SH3
            homology
FEATURE
89-137      #domain SH3 homology #label SH3\
150-246     #domain SH2 homology #label SH2\
269-527     #domain protein kinase homology #label KIN
SUMMARY
            #length 527 #molecular-weight 61239 #checksum 7997
Query Match 14.9%; Score 82; DB 2; Length 527;
Best Local Similarity 44.4%; Pred. No. 2.15e+00;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;
Db 242 LRVPGVGLMGSCPLPATAGFSYKWEIDP 268
QY 24 VRYKISAGSCPLSTAGPSYVKFQDNP 50

RESULT 12
ENTRY
TITLE      G70447      #type complete
            flagellar L-ring protein FlgH - Aquifex aeolicus
ORGANISM    #formal_name Aquifex aeolicus
DATE        08-May-1998 #sequence_revision 08-May-1998 #text_change
            08-May-1998
ACCESSIONS  G70447
REFERENCE    A70300
            #authors
            Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
            Lennox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;
            Keller, M.; Aulay, M.; Huber, R.; Feldman, R.A.; Short,
            J.M.; Olson, G.J.; Swanson, R.V.
            #journal Nature (1998) 392:353-358
            #title The complete genome of the hyperthermophilic bacterium
            #title Aquifex aeolicus.
            #cross-references MUID:98196666
            #accession G70447
            ##status preliminary; nucleic acid sequence not shown;
            #molecule_type DNA
            ##residues 1-216 #label AQF
            ##cross-references GB:AE000752; NID:g2984021; PID:g2984027; GB:AE000657
            ##experimental_source strain VP5
GENETICS
            #gene flgH
            #length 216 #molecular-weight 23767 #checksum 9605
SUMMARY
Query Match 14.7%; Score 81; DB 2; Length 216;
Best Local Similarity 27.7%; Pred. No. 2.96e+00;
Matches 18; Conservative 25; Mismatches 16; Indels 6; Gaps 5;
Db 93 SSFPGVHPATLK-NLGAGKSSPASKGS--KFOQSGVLTTTL-AGRVVKVF-PNGTMLV 147
QY 15 TPFLS-ESNSVRKYISAGSCPLSTAGPSYVKFQDNPVGSGTFSAGLHRLVDFDPSTGALV 73
Db 148 EAKKY 152
QY 74 DSKSY 78

RESULT 13
ENTRY
TITLE      A55631      #type complete
            protein-tyrosine kinase (EC 2.7.1.112) rlk - mouse
ALTERNATE_NAMES
ORGANISM    resting lymphocyte kinase
DATE        #formal_name Mus sp. #common_name mouse
            23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change
            17-Mar-1999
ACCESSIONS  A55631
REFERENCE    A55631
            #authors
            Hu, Q.; Davidson, D.; Schwartzberg, P.L.; Macchiarini, F.;
            Lenardo, M.J.; Bluestone, J.A.; Matis, L.A.
            #journal J. Biol. Chem. (1995) 270:1928-1934
            #title Identification of rlk, a novel protein tyrosine kinase with
            #title predominant expression in the T cell lineage.
            #cross-references MUID:95130578
            #accession A55631
            ##status preliminary
            ##molecule_type mRNA
            ##residues 1-527 #label HUA
            ##cross-references GB:L35268; NID:g623442; PID:g623443
            #superfamily protein-tyrosine kinase tec; pleckstrin repeat
            #homology; protein kinase homology; SH2 homology; SH3
            homology
            phosphotransferase
KEYWORDS
FEATURE
89-137      #domain SH3 homology #label SH3\
150-246     #domain SH2 homology #label SH2\
269-527     #domain protein kinase homology #label KIN\
277-285     #region protein kinase ATP-binding motif
SUMMARY
            #length 527 #molecular-weight 61108 #checksum 785
Query Match 14.7%; Score 81; DB 2; Length 527;
```

[illegible]

	CC	TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).	
	CC	- - - - -	
	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to licensee@isb-sib.ch).	
	CC	- - - - -	
	DR	EMBL; D10330; G221705; -. RNA-DIRECTED RNA POLYMERASE; HELICASE;	
	KW	POLYPROTEIN; TRANSFERASE;	
	FT	NP-BINDING.	
	KW	NP_BIND 975 982 ATP (POTENTIAL).	
	SQ	SEQUENCE 1693 AA; 185215 MW; FFCB786D CRC32;	
		Query Match 18.1%; Score 100; DB 1; Length 1693; Best Local Similarity 26.9%; Pred. No. 9.43e-04; Matches 18; Conservative 22; Mismatches 24; Indels 3; Gaps 2;	
Df	Db	572 FRTSFVGVLEANGPPRYNLSFDASQTWAAGPFSLTYAASAAGLVRRVAACGLDHRAV 631 : : : : : : : : : : : : : : : : + FGQYGVTPLSESNVRKYKSIAGSCPLSTAGSYVKFQDNPVGSQT-F-SAGHLR-V 64	
Dy	Qy	632 FAGCVSP 638 : : 65 FDPSTA 71	
	RESULT 3	STANDARD; PRT; 457 AA. ID EMB8_PIGCL STANDARD; PRT; 457 AA. AC Q40863; DT DT 01-NOV-1997 (REL. 35, CREATED) DI 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE) DL 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE) DE LATE EMBRYOGENESIS ABUNDANT PROTEIN EMB8. DS EMB8. GN PICEA GLAUCA (WHITE SPRUCE). OS EUKARYOTA; VIRIDIPLANIAE; STREPTOPHYTA; TRACHEOPHYTA; OC EUPHYLLIPHYTES; SPERMATOPHYTA; CONIFEROPSIDA; PINACEAE; RN [1] RP SEQUENCE FROM N.A. RA DONG J.Z., DUNSTAN D.I.; RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS. RC - - SIMILARITY: BELONGS TO THE UPF0017 FAMILY. CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to licensee@isb-sib.ch).	
	CC	- - - - -	
	DR	EMBL; LA7118; GI350545; -. RNA-DIRECTED RNA POLYMERASE; HELICASE;	
	KW	POLYPROTEIN; TRANSFERASE;	
	FT	NP-BINDING.	
	KW	NP_BIND 975 982 ATP (POTENTIAL).	
	SQ	SEQUENCE 457 AA; 51019 MW; AE7CB4CD CRC32;	
		Query Match 15.6%; Score 86; DB 1; Length 457; Best Local Similarity 34.0%; Pred. No. 1.92e-01; Matches 17; Conservative 11; Mismatches 20; Indels 2; Gaps 2;	
Df	Db	232 LGANILRVYLGEVAGNCPLSGAVSLCPNF-LNLVIADDFHKGFGNNVDY 280 : : : : : : : : : : : : : : : : + LSESNSVRKYKSIAGSCPLSTAGSYVKFQDNPVGSQTFSSAGHLR-VFD 66	
Dy	Qy	18 LSESNSVRKYKSIAGSCPLSTAGSYVKFQDNPVGSQTFSSAGHLR-VFD 66	
	RESULT 4	STANDARD; PRT; 1693 AA. ID POLN_HEVB0 STANDARD; PRT; 1693 AA. AC R29324; DT DT 01-DEC-1992 (REL. 24, CREATED) DL 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)	


```

CC -----
DR EMBL; M80381; G329998; -
KW POLYPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; HELICASE;
KW ATP-BINDING.
FT NP_BIND 975 982 ATP (POTENTIAL).
SQ SEQUENCE 1693 AA; 185149 MW; FBCA2483 CRC32;

Query Match 15.6%; Score 86; DB 1; Length 1693;
Best Local Similarity 25.4%; Pred. No. 1.92e-01;
Matches 17; Conservative 22; Mismatches 25; Indels 3; Gaps 3;

Db 572 FRTSFDGAVLTNGPERNLNLSFDSQSTMAAGPSLTYAASAGLEVYVVAAGLDHRAV 631
QY 8 FCGYVQTPFLSESNSVRYKISAGSCPLSTAGSPSYVKFQDNPVGSQT-F-SAGLHLR-V 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 632 FAPGVSP 638
QY 65 FDPSTGA 71
| : : :
| : : :

RESULT 6
ID Y411 RHISN STANDARD; PRT; 703 AA.
AC P55492;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 76.2 KD PROTEIN Y411.
GN Y411.
OS RHIZOBIUM SP. (STRAIN NGR234).
OC PLASMID SYN PNR234A.
CC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHIZOBIACEAE GROUP;
CC RHIZOBIACEAE; RHIZOBIUM.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97305956.
RA FREIBERG C.A., FELLAY R., BAIRDOCH A., BROUGHTON W.J., ROSENTHAL A.,
RA PERRET X.;
RT Molecular basis of symbiosis between Rhizobium and legumes.*;
RL NATURE 387:394-401(1997).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: NONE OBVIOUS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AE000078; G2182446; -
KW HYPOTHETICAL PROTEIN; PLASMID; TRANSMEMBRANE.
FT TRANSMEM 23 43 POTENTIAL.
FT TRANSMEM 69 89 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 250 270 POTENTIAL.
FT TRANSMEM 357 377 POTENTIAL.
FT TRANSMEM 432 452 POTENTIAL.
FT TRANSMEM 644 664 POTENTIAL.
SQ SEQUENCE 703 AA; 76183 MW; A2BA53CE CRC32;

Query Match 15.4%; Score 85; DB 1; Length 703;
Best Local Similarity 27.0%; Pred. No. 2.75e-01;
Matches 17; Conservative 17; Mismatches 28; Indels 1; Gaps 1;

Db 224 SPFADTDSGTSTAAATSLFGAGAGATRPPEWLAFAAALASGAPFGLSPRVATATATSYA 283
QY 15 TPFL-SESNSVRYKISAGSCPLSTAGSPSYVKFQDNPVGSQTF-SAGLHLRVPDSTGALV 73
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 284 KAK 286
QY 74 DSK 76
| :
| :

```

```

RESULT 7
ID CFAC_ECOLI STANDARD; PRT; 869 AA.
AC P25733;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE CFA/I FIMBRIAL SUBUNIT C PRECURSOR (COLONISATION FACTOR ANTIGEN I
DE SUBUNIT C).
GN CFAC.
OS ESCHERICHIA COLI.
OG PLASMID NTF513.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-ENTEROTOXIGENIC;
RX MEDLINE; 89330163.
RA HAMERS A.M., PEL H.J., WILLSHAW G.A., KUSTERS J.G.,
RA VAN DER ZEIJST B.A.M., GAASTRA W.;
RT *The nucleotide sequence of the first two genes of the CFA/I fimbrial
RT operon of human enterotoxigenic Escherichia coli.*;
RL MICROB. PATHOG. 6:297-309(1989).
[2]
RN SEQUENCE FROM N.A.
RP MEDLINE; 92329981.
RA JORDI B.J.A.M., WILLSHAW G.A., VAN DER ZEIJST B.A.M., GAASTRA W.;
RT *The complete nucleotide sequence of region 1 of the CFA/I fimbrial
RT operon of human enterotoxigenic Escherichia coli.*;
RL DNA SEQ. 2:257-263(1992).
CC -!- FUNCTION: MAY SERVE AS ANCHOR FOR THE FIMBRIAE IN THE OUTER
CC MEMBRANE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M55661; G145510; -
KW ANTIGEN; SIGNAL; FIMBRIA; OUTER MEMBRANE; PLASMID.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 869 CFA/I FIMBRIAL SUBUNIT C.
SQ SEQUENCE 869 AA; 97830 MW; 7AF76347 CRC32;

Query Match 15.4%; Score 85; DB 1; Length 869;
Best Local Similarity 32.4%; Pred. No. 2.75e-01;
Matches 11; Conservative 14; Mismatches 8; Indels 1; Gaps 1;

Db 158 AFIOQTINLSDSGYKRLISGNSALGITDTSY 191
QY 11 GYVQTPFLSESNSVRYK-ISIAGSCPLSTAGPSY 43
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
ID YX45_MYCTU STANDARD; PRT; 611 AA.
AC Q10830;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 67.2 KD PROTEIN CY274.45C.
GN MYC274.45C OR MYC338.01C.
OS MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIACEAE;
OC ACTINOMYCETACEAE; CORYNEBACTERIACEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-H37RV;
RA CONNOR R., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SIMILARITY: BELONGS TO THE N-ACYL-D-AMINO-ACID DEACYLASE FAMILY.

```



```
FT CONFLICT 45 45 H -> R (IN REF. 2).
SQ SEQUENCE 527 AA: 61239 MW: 8DF019E3 CRC32:

Query Match
Best Local Similarity 44.4%; Pred. No. 7.96e-01;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 242 LRYVGLMGSLCPATAGFSYKWEIDP 268
QY 24 VRYKISAGSLPLSTAGPSYVAFQDNP 50

RESULT 11
ID TXK_MOUSE STANDARD: PRT: 527 AA.
AC P42682;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TYROSINE-PROTEIN KINASE TXK (EC 2.7.1.112) (PTK-RL-18) (RESTING
DE LYMPHOCYTE KINASE).
GN TXK OR RLK.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE-THYMUS;
RX MEDLINE: 96059536.
RA HAIRE R.N., LITMAN G.W.;
RT "The murine form of TXK, a novel TEC kinase expressed in thymus maps
RT to chromosome 5."
RL MAMM. GENOME 6:476-480(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-FVB/N; TISSUE-THYMUS;
RA SOMMERS C.L., HUANG K., GRINBERG A., CHARLICK D.A., KOZAK C.A.,
RA LOVE P.E.;
RL SUBMITTED (JAN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6; TISSUE-LIVER;
RA HIGASHITSUJI H., NONOGUCHI K., ARII S., FURUTANI M., KANEKO Y.,
RA NAKAYAMA H., FUJITA J.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE-THYMUS;
RX MEDLINE: 95130578.
RA HU Q., DAVIDSON D., SCHWARTZBERG P.L., MACCHIARINI F.,
RA LENARDO M.J., BLUESTONE J.A., MATIS L.A.;
RT "Identification of Rlk, a novel protein tyrosine kinase with
RT predominant expression in the T cell lineage."
RL J. BIOL. CHEM. 270:1928-1934(1995).
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN EARLY THYMOCYTES, T CELLS AND
CC MAST CELLS.
CC -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN. BELONGS TO THE BTK SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: U16145; G562125; -.
EMBL: U19607; G643065; -.
-----
```

```
DR EMBL: D43963; G604884; -.
DR MGD; MGI:102960; TXK.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS50001; SH2; 1.
DR PROSITE: PS50002; SH3; 1.
DR PFAM: PF00017; SH2; 1.
DR PFAM: PF00018; SH3; 1.
DR PFAM: PF00069; pkinase; 1.
DR HSP; Q06187; IAWW.
KW TRANSFERASE; TYROSINE-PROTEIN KINASE; ATP-BINDING; SH2 DOMAIN;
KW SH3 DOMAIN; PHOSPHORYLATION.
FT DOMAIN 14 20 POLY-CYS.
FT DOMAIN 82 142 SH3.
FT DOMAIN 150 246 SH2.
FT NP_BIND 271 285 PROTEIN KINASE.
FT BINDING 299 299 ATP (BY SIMILARITY).
FT ACT_SITE 390 390 ATP (BY SIMILARITY).
FT MOD_RES 420 420 BY SIMILARITY.
FT CONFLICT 3 4 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CONFLICT 6 6 LS -> SF (IN REF. 3).
FT CONFLICT 272 272 Y -> D (IN REF. 3).
FT CONFLICT 497 497 A -> T (IN REF. 3).
FT CONFLICT 497 497 R -> S (IN REF. 3).
SQ SEQUENCE 527 AA: 61108 MW: 5B39DA78 CRC32:

Query Match 14.7%; Score 81; DB 1; Length 527;
Best Local Similarity 44.4%; Pred. No. 1.13e-00;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

Db 242 LRYVGLMGSLCPATAGFSYKWEIDP 268
QY 24 VRYKISAGSLPLSTAGPSYVAFQDNP 50

RESULT 12
ID YCB2_PSEDE STANDARD: PRT: 171 AA.
AC P29944;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 19.0 KD PROTEIN IN COBS 5' REGION (ORF2).
OS PSEUDOMONAS DENITRIFICANS.
OC BACTERIA; PROTEOBACTERIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92011364.
RA CAMERON B., GUILHOT C., BLANCHE F., CAUCHOIS L., ROUYEZ M.-C.,
RA RIGAUD S., LEVY-SCHIL S., CROUZET J.;
RT "Genetic and sequence analyses of a Pseudomonas denitrificans DNA
RT fragment containing two cob genes."
RJ J. BACTERIOL. 173:6058-6065(1991).
CC -!- SIMILARITY: CONTAINS A DNAJ-LIKE DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: M62869; G151167; -.
PIR: B38162; B38162.
DR PROSITE: PS00636; DNAJ_1; FALSE_NEG.
DR PROSITE: PS50076; DNAJ_2; 1.
DR PFAM: PF00226; DnaJ; 1.
DR HSP; P08622; 1XBL.
DR HYPOTHETICAL PROTEIN.
SQ SEQUENCE 171 AA: 18973 MW: 5DDDD21D5 CRC32:
```

```

Query Match      14.5%: Score 80; DB 1; Length 171;
Best Local Similarity 37.5%: Pred. No. 1.59e+00;
Matches 24; Conservative 11; Mismatches 25; Indels 4; Gaps 4;

Db 24 EYNGKGYFFSGLSSEVAVRQKEAITHRPTWTVGVN-KNAKNQPTOSQTRSGSAGAAQAR 82
QY 7 EFGQGYVOTPFLESNSVRY-KISAGSCPLSTAGPSYKRVQDNVPVGSQT-F-SAGLHLR 63

Db 83 MRDP 86
QY 64 VFDP 67

RESULT 13
ID CV42.TRYBB STANDARD; PRT; 572 AA.
AC Q9396;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE RECEPTOR-TYPE ADENYLATE CYCLASE GRESAG 4.2 (EC 4.6.1.1) (ATP
DE PYROPHOSPHATE-LYASE) (ADENYL CYCLASE) (FRAGMENT).
GN GRESAG 4.2.
OS TRYPANOSOMA BRUCEI BRUCEI.
OC EUKARYOTA; EUGLENOZOA; KINETOPLASTIDA; TRYPANOSOMATIDAE; TRYPANOSOMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EATRO 1125;
RX MEDLINE: 91218809.
RA ALEXANDRE S., PAINDAVIONE P., TEBABI P., PAYS A., HALLEUX S.,
RA STEINERT M., PAYS E.;
RT "Differential expression of a family of putative adenylate/guanylate
RT cyclase genes in Trypanosoma brucei.";
RL MOL. BIOCHEM. PARASITOL. 43:279-288(1990).
CC -1- FUNCTION: COULD ACT AS A RECEPTOR FOR A UNKNOWN LIGAND.
CC -1- CATALYTIC ACTIVITY: ATP -> 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC -1- SIMILARITY: BELONGS TO CLASS-3 OF ADENYL CYCLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52120; E30923;
DR PFAM: PF00211; guanylate_cyc; 1.
DR HSP: Q02846; IAWL.
KW LYASE; CAMP SYNTHESIS; TRANSMEMBRANE; RECEPTOR; GLYCOPROTEIN.
FT NON_TER 1
FT DOMAIN <1 225 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 226 251 POTENTIAL.
FT DOMAIN 252 572 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 10 10 POTENTIAL.
FT CARBOHYD 77 77 POTENTIAL.
FT CARBOHYD 152 152 POTENTIAL.
SQ SEQUENCE 572 AA; 63683 MW; CE3202BF CRC32;

Query Match      14.5%: Score 80; DB 1; Length 572;
Best Local Similarity 30.0%: Pred. No. 1.59e+00;
Matches 15; Conservative 16; Mismatches 17; Indels 2; Gaps 2;

Db 442 LSTAERSQFDVTLGGVPLRGVSEPVYQLNAVPGRSF-AELRLDRLVD 490
QY 18 LSESNSVRYKISAGSCPLSTAGPSYKRVQDNVPVGSQTFSGAGLHL-RVFD 66

RESULT 14
ID NAB3_YEAST STANDARD; PRT; 802 AA.
AC P38996;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)

```

```

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE NUCLEAR POLYADENYLATED RNA-BINDING PROTEIN NAB3.
GN NAB3 OR YPL190C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RA WILSON S.M., OBERDORF A.M., DATAR K.V., SWEDLOW J.R., PADDY M.R.,
RA SWANSON M.S.;
RL SUBMITTED (JAN-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA RIEGER M., MUELLER-AUER S., SCHAEFER M.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNP).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U05314; G476220;
DR EMBL: Z73546; E246917;
DR PIR: S48529; S48529.
DR SGD; L0001228; NAB3.
DR PFAM: PF00076; rtm; 1.
KW RNA-BINDING; NUCLEAR PROTEIN.
FT DOMAIN 68 71 POLY-GLU.
FT DOMAIN 87 93 POLY-GLU.
FT DOMAIN 101 106 POLY-ASP.
FT DOMAIN 108 115 POLY-GLU.
FT DOMAIN 116 127 POLY-ASP.
FT DOMAIN 128 137 POLY-GLU.
FT DOMAIN 603 608 POLY-GLN.
FT DOMAIN 644 648 POLY-PRO.
FT DOMAIN 698 703 POLY-GLN.
FT DOMAIN 723 728 POLY-GLN.
FT DOMAIN 765 768 POLY-PRO.
FT DOMAIN 769 784 POLY-GLN.
SQ SEQUENCE 802 AA; 90438 MW; FB180EDB CRC32;

Query Match      14.5%: Score 80; DB 1; Length 802;
Best Local Similarity 32.6%: Pred. No. 1.59e+00;
Matches 14; Conservative 10; Mismatches 16; Indels 3; Gaps 3;

Db 622 QGYGSOPPIPMNOSYG-RYOTSIPPPPPQOQIPOGYGRYQAGP 663
QY 10 QGY-VQTPF-LSESNSVRYKISAGSCPLSTAGPSYKRVQDNVP 50

RESULT 15
ID YB79_YEAST STANDARD; PRT; 954 AA.
AC P38138;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PUTATIVE FAMILY 31 GLUCOSIDASE IN PCS60-ABD1 INTERGENIC REGION
DE (EC 3.2.1.-).
GN YBR290C OR YBR1526.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RA DUBOIS E., EL BAKKOURY M., GLANSDORFF N., MESSENGUY F., PIERARD A.,
RA SCHERENS B., VIERENDELS F.;
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.

```

